

Covert Embodied Choice: Decision-Making and the Limits of Privacy Under Biometric Surveillance

Jeremy Gordon
UC Berkeley
jrgordon@berkeley.edu

John Chuang
UC Berkeley
john.chuang@berkeley.edu

Max Curran
UC Berkeley
mtcurran@berkeley.edu

Coye Cheshire
UC Berkeley
coye@berkeley.edu



Figure 1: Left: participant in VR experiment. Right: participant field of view during card selection.

ABSTRACT

Algorithms engineered to leverage rich behavioral and biometric data to predict individual attributes and actions continue to permeate public and private life. A fundamental risk may emerge from misconceptions about the sensitivity of such data, as well as the agency of individuals to protect their privacy when fine-grained (and possibly involuntary) behavior is tracked. In this work, we examine how individuals adjust their behavior when incentivized to avoid the algorithmic prediction of their intent. We present results from a virtual reality task in which gaze, movement, and other physiological signals are tracked. Participants are asked to decide which card to select without an algorithmic adversary anticipating their choice. We find that while participants use a variety of strategies, data collected remains highly predictive of choice (80% accuracy). Additionally, a significant portion of participants became

more predictable despite efforts to obfuscate, possibly indicating mistaken priors about the dynamics of algorithmic prediction.

CCS CONCEPTS

• **Human-centered computing** → *Virtual reality*; • **Social and professional topics** → **Surveillance**;

KEYWORDS

biometrics, prediction, privacy, virtual reality, surveillance

ACM Reference Format:

Jeremy Gordon, Max Curran, John Chuang, and Coye Cheshire. 2021. Covert Embodied Choice: Decision-Making and the Limits of Privacy Under Biometric Surveillance. In *CHI Conference on Human Factors in Computing Systems (CHI '21)*, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3411764.3445309>

1 INTRODUCTION

In a world in which sensing devices permeate both public and private domains, consumers [17], renters [13], voters [9], and decision-makers of all kinds are now subject to unprecedented levels of surveillance during day-to-day life: from the tracking of search queries to face and ‘emotion recognition’ [8], to geospatial location and physiology [4]. The data produced by sensors and software



This work is licensed under a Creative Commons Attribution International 4.0 License.

CHI '21, May 8–13, 2021, Yokohama, Japan

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8096-6/21/05.

<https://doi.org/10.1145/3411764.3445309>

capturing these signals are now and will continue to be analyzed by increasingly advanced statistical techniques and algorithms with a specific interest in inferring internal states that might predict the future actions of individuals. Constituting what Zuboff terms the ‘Big Other’, this constellation of tracking infrastructure uses opaque mechanisms to detect and control behavior, and ultimately poses a threat to both individual privacy and democratic norms at large [38].

According to Nissenbaum’s contextual integrity heuristic, privacy is best understood as a set of appropriate information flows subject to contextual norms dependent on parameters such as the data subject, parties, information type, and principles of transmission [26]. As proposed by Sedenberg et al. 2017, the privacy threats posed by remote biometric sensing can be seen as shifts or breaches of these contextual norms: “...when biosensed data like emotions or internal physiological states are systematically recorded and analyzed, all signals become magnified beyond their original natural public scope” [34]. Indeed, remote sensing systems of this kind have already generated real-world contexts within which actors may hope to predict internal beliefs or preferences from biometric data, and for which the obfuscation of these beliefs may be beneficial to the individual. One such example is individually targeted price discrimination in a physical retail setting where video feeds may afford gaze tracking (see commercial use of “Smart-shelf” technology reported in 2013[7]).

Thanks to journalistic and public service efforts, an awareness of these threats is growing, but individuals still hold significant misconceptions about the sensitivity of information that can be inferred from these types of data [23], and what agency they have in protecting their privacy.

In this work, we designed a Virtual Reality (VR)-based behavioral experiment in which participants completed an iterative binary decision task. In later trials, participants were informed of an adversary that is tracking their behavior and attempting to predict each choice. Questionnaire responses, post-task interviews and quantitative analyses of biometric data indicate that participants used a variety of behavioral strategies and felt confident in their agency in avoiding prediction. While some participants adjusted their behavior in ways that reduced prediction accuracy, a strong majority of trials could still be successfully classified based on biometric data despite these efforts.

1.1 Background

1.1.1 Embodied Cognition and Decisions. A valuable perspective on the privacy of decisions in a tracked environment comes from a multidisciplinary literature which integrates embodied, enactive, extended, and embedded accounts of cognition (referred to as ‘4E Cognition’). In a 2015 paper, Lepora and Pezzulo introduced the Embodied Choice framework (EC) which recasts decision-making as a dynamic and inherently active process rather than a sequential perception-decision-execution cycle [18]. In a visual decision task, the authors tracked the correlates of active consideration via movements of a mouse towards a target. These ideas are inspired by and consistent with Active Inference, a formal framework modeling situated action in which agents minimize their prediction error by entraining available sensors in such a way as to disambiguate

competing perceptual hypotheses [35]. In this framework, percepts can be seen as hypotheses (e.g. *the dog is hungry*, or *the customer is planning to buy some flowers*), and the ocular motor outputs that produce saccades (eye movements) are seen as experiments aiming to confirm or deny prior beliefs [12].

In another behavioral experiment looking at embodied decision-making, Beilock and Holt found that the visual perception of stimuli can recruit the motor system into action simulation (the production of potentially detectable micro-motor outputs) which can influence affective judgments [6]. Together, research within this school of thought encourages us to reconsider traditional notions of passive perception followed by active decisions, and instead see decision-making as a continuous and dynamic process. In this light, sensed biometric data may serve as a correlate to deliberative processes. The privacy implications stemming from this reasoning have not yet been sufficiently investigated.

1.1.2 Eye Tracking, Virtual Reality, and Privacy. Eye-tracking technology is far from new, but the increasing prevalence of video capture systems and improved algorithms for accurately inferring gaze direction [19] make it especially relevant to conversations around modern algorithmic surveillance. Further, the breadth of application areas being explored for eye-tracking—as a novel UI selection affordance [30], a method to augment user experience via attentional awareness [37], and an opportunity to gain computational efficiencies via foveated rendering [27], among many others—reinforces the belief of some researchers that eye tracking technology will soon be ubiquitous [20].

Beyond eye-tracking, a critical body of literature has explored the deleterious effects of various regimes of biometric surveillance [22, 31], including works specifically highlighting the dangers when these technologies are targeted at marginalized groups [21], and the unique effects of systems purporting to monitor internal states such as affect [8]. The documentation of surveillance harms offered by these scholars, among others, provided a key motivation for the present work.

Scholars in HCI have offered a variety of perspectives on privacy concerns relating to the use of consumer remote sensing technologies such as Internet of things (IoT) products like smart speakers, and augmented reality (AR) devices. Denning et al. explored the perceptions and privacy concerns of bystanders to the use of glasses-style AR devices which may record video of people and surroundings [10]. In a 2014 review, Roesner et al. identified several key challenges and risks of AR systems, including adversarial attacks to both input and output channels, and theft or misuse of sensor data [33]. Recent work by Ahmad et al. raises concerns over privacy ambiguity imposed by ‘always-on’ IoT devices, and introduces the concept of tangible privacy—design features that allow bystanders to clearly assess the state of a device’s sensors [1].

We chose to implement our study using a VR-based laboratory precisely because of the ease of collecting not only gaze targets, but a rich stream of sensor data allowing us to measure our participants’ in-task behavior at a fine level of granularity. Indeed, the potential to develop machine learning-based predictive models trained on VR biometric data, some of it imperceptible to human observers, has not gone unnoticed by businesses and entrepreneurs [24]. The ethical questions raised by this new domain, which is quickly developing

on interrelated but discrete paths in the private sector and academia, are diverse and poignant. A first of its kind conference was held in 2018 proposing to construct a “VR bill of rights” [15], and an opinion article by Jeremy Bailenson in the same year highlighted the capabilities and dangers of biometric data collected in VR [3].

While immersive VR clearly comes with a host of potential privacy concerns, our primary interest centers around VR as a research tool. The VR-based setup allows us to probe the sensitivity of collected data, as well as the psychology influencing the perception of surveillance, in both in-home and public settings. Ultimately, we see the virtual laboratory as offering researchers a controllable, replicable setting enabling the study of human behavior in a real world increasingly subject to surveillance.

1.2 Research Questions

Though biometric surveillance has attracted extensive scholarship, few quantitative behavioral studies have shed light on the expectations, assumptions, and range of behavioral responses employed by individuals hoping to evade prediction by an adversarial system. In this work, we aim to explore the following questions:

- (1) RQ1: Under what conditions might biometric signals such as motor outputs, eye movements, and electrodermal activity expose sensitive information relating to beliefs and immediate choice intentions?
- (2) RQ2: How effectively can commonly used machine learning models predict choice intention given behavioral data during decision-making?
- (3) RQ3: What strategies are used by participants when instructed to make unpredictable decisions in a simple tracked setting?
- (4) RQ4: How effective are the employed strategies at maintaining an individual’s privacy of intent?

We note that as per RQ2, we seek to understand the approximate performance of minimally tuned off-the-shelf tools, rather than to estimate an upper bound on prediction accuracy. The development of novel algorithms that might improve performance on this prediction task were not among the objectives of this work (see Section 4.4 for discussion).

2 METHODS

2.1 Task Design

In order to study human behaviors related to the protection of private intent, we developed a game-based experiment in which participants were sufficiently motivated to avoid predictability while behaving in a physical environment. Specifically, we defined the following desiderata to constrain experimental design:

- (1) **Covert imperative.** Participants should be sufficiently incentivized to obscure their intent.
- (2) **Non-trivial choice-salience.** Decisions made in the task must influence a relevant outcome (e.g. compensation); participants must consider an extrinsic value of their choice separate from masking intent.
- (3) **Embodied.** Choices should be enacted by gross motor outputs as opposed to verbal reports to allow for action simulation or other micro-motor outputs to be detectable.

2.2 Covert Embodied Choice Task

Based on these criteria, we designed a virtual reality task (inspired by the popular game *Set*) as follows.

In each trial, the participant is presented with two cards in a private ‘hand’ facing them, and two cards flat on a table (see Figure 2). During the 10-second decision phase, they must choose one of the two single cards on the table. Then, during the 3-second selection phase, they use the controller to bring the chosen card into their hand to complete a trial. A trial is successful if the resulting three cards in the hand form a complete ‘match’. A successful match is defined as three cards for which both card attributes: count and shape are either all the same or all different. A trial is completed either upon the participant’s releasing their chosen card into their hand, or when the selection timer elapses. A new trial is started after completion of the last, with a new deal of two private and two table cards.

See Section 2.4 for a more detailed description of the task and experimental protocol.

2.3 Participants

Study participants were recruited via e-mail through an on-campus experimental lab at a public university in the United States. Participants were screened to minimize the risk of adverse effects from the use of VR: participants indicating dizziness or nausea during prior uses of VR, epilepsy, a history of seizures, or that they are prone to motion sickness, were not considered for the full study.

To support the use of parametric tests with our mixed between-subjects and within-subjects design, we selected a target sample size of 35 subjects per group (70 total). Since machine-learning-based classification accuracy was a primary dependent variable, it was critical to ensure both that (1) the training set contained sufficient data with respect to the dimensionality of the feature-space and complexity of the model to effectively generalize, and (2) that the test set was large enough to accurately estimate model performance. Despite stopping data collection early due to the COVID-19 pandemic (with 57 completions), our resultant dataset included 1200 non-practice subject-trials which we deemed sufficient for our planned analysis.

Two participants were dropped after they indicated in the post-experiment interview that they misunderstood the instructions and believed the task was to enable the adversary to predict their intent (PP 004), or provide training labels to the adversary (PP 041).

Participants were randomly placed into one of the two study conditions: early adversary (N=27), or delayed adversary (N=28). The two study conditions operated identically except for the trial number at which the adversary info screen appeared (see 2.4), allowing us to control for potential learning effects. Participants had a mean age of 20.4 ± 4.6 , 34 identified themselves as female, and 21 as male. 47.3% reported their race as Asian, 34.5% white, 1.8% American Indian or Alaska Native, 1.8% Native Hawaiian or Pacific Islander, and 14.5% Other, including Asian & white, and Middle Eastern. 16.4% of participants identified as Spanish, Hispanic, or Latino. All participants were undergraduate students, graduate students or staff at the university where the study was conducted.

2.4 Experimental Procedure

After providing consent, each participant went through a self-paced slide-based tutorial using the experiment computer. The tutorial explained the pattern matching task, and provided quizzes to ensure the participant had understood how to identify a match consistent with the rules of the game. The tutorial also detailed how a final success score would be calculated, and the monetary bonus that would be provided depending on this score. Bonuses were calculated based on the percent of successful rounds (a correct match for pre-adversary rounds, and both a correct match and the avoidance of prediction in the adversarial rounds), however information about the adversary was not presented until later. All participants received the base compensation (\$20) for their participation, with performance-based rewards contributing an additional 0-25% of the base level. At the end of the tutorial the experimenter gave the participant an opportunity to ask any clarifying questions.

Next, the experimenter fitted the participant with an Empatica E4 skin conductance and heart rate monitor, and a Vive Pro Eye VR headset. The participant completes a brief exercise in which they fixate on targets that appear on the screen in sequence to calibrate the in-unit eye tracking system. Once complete, the participant was asked to use the trigger on the controller to begin the first of four (unscored) practice rounds in the experiment. The practice rounds simulate a normal trial, and allow the participant to become comfortable with the mechanics of the interaction in VR. Practice rounds, like regular rounds, had a 10-second decision timer, and 3-second selection timer, so that participants learned how to make their selection in the allotted time.

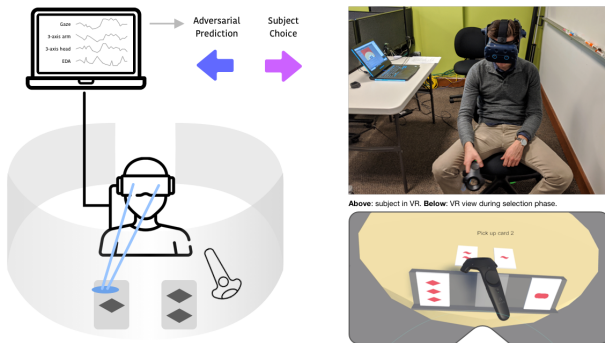


Figure 2: Left: schematic of subject during a trial. Top-right: participant in VR. Bottom-right: participant’s field of view during card selection.

After completing all four practice rounds, the participants were asked to confirm they understood the task and the interaction, and if so, they used the controller trigger to begin the main experiment.

After three trials (early condition) or 22 trials (delayed condition, midway point) the participant was presented with a screen informing them that the task would be changing and that in all remaining trials, an adversary would be “tracking their behavior during the decision phase” and attempting to predict which card they would choose. They were also informed that success on each following trial would be determined by *both* selecting a correct match *and*

avoiding their choice being predicted by the adversary¹. After the participant reviewed this information, the experimenter verbally asked whether the information was clear and prompted for any final questions before instructing the participant to continue with the experiment when ready.

All participants completed all 44 non-practice trials. Upon completion of the last trial, a screen appeared summarizing the participant’s performance across all trials (counts of correct matches, and successful avoidance of prediction), and displaying their final success score and bonus compensation earned.

After removing the VR headset, the participant was asked to respond to a short questionnaire and structured interview covering their beliefs about the biometric data that was used, strategies they employed, and confidence in their efficacy². Once complete, a debrief form and media release form were signed, and finally the participant was provided their incentive payment and told they may exit the lab.

The recruitment process and study protocol was approved by the local ethics review board.

2.5 Data Structure, Cleaning, and Analysis

The VR experiment was implemented in C# using the Unity Editor (version 2019.3.0f6), Steam VR (version 1.9.16), and Steam VR Unity Plugin (version 2.5.0). Telemetry was recorded in the update loop and included 3-dimensional position and rotation vectors for the HMD (head motion), and controller (arm motion). Two types of gaze data were collected. Fixations were recorded using the Vive Pro Eye SRanipal SDK (version 1.1.0.1) which infers in-scene gaze target objects via the Tobii Gaze-to-Object-Mapping algorithm³. This data type was converted into a series of fixation records tracking start and stop timestamps, as well as the scene element fixated. Secondly, raw gaze data was recorded in the update loop providing a gaze origin and direction vector in world coordinates, as well as an approximate vergence distance allowing tracking of the implied (x, y, z) coordinates of raw gaze target.

Physiology data including heart rate, inter-beat interval (IBI), and electrodermal activity, was extracted from data files produced by the E4. We extracted EDA spike timestamps using peakutils [25], and computed heart rate variability (HRV) based on IBI records using the *HRV* Python library [5]. All data was aligned and registered by comparing synchronized device timestamps and trial-by-trial metadata recorded by the Unity experiment. All experiment code is available in an MIT licensed repository: https://github.com/onejgordon/cec_vr.

¹During experimentation, the adversary was programmed to make predictions based on a simple gaze-based heuristic trivial to compute in real-time. Because only an aggregate success score was provided to participants after all trials were complete, the specific mechanism of prediction could not influence participant behavior.

²The questionnaire and interview were intended as exploratory methods to capture unexpected responses and hear from participants in their own words. While the experimental study and quantitative results are the primary contributions of this work, the qualitative data allowed for more nuanced interpretation and set up possible focus areas for future, more in-depth investigation.

³For details on Tobii’s G2OM algorithm, see <https://vr.tobii.com/sdk/technology/tobii-g2om/>

2.6 Predictive Model

To assess participants' effectiveness at protecting privacy of intent, we trained a discriminative machine learning model to predict a participant's ultimate choice of card. We excluded trials where the participant failed to choose either card in the allotted time, so the model was required only to correctly predict a binary label corresponding to *choose left* or *choose right*. Data available to the predictive model included all physiological and behavioral data collected during the 10-second decision phase, which was clearly demarcated to participants by a text alert and changing background color (see section 2.6.2 for details). Data captured during the selection phase, such as participants' arm motion to pick up a card, was not available to the algorithm.

We conducted experiments within two related prediction paradigms both consistent with plausible real-world settings but presenting unique challenges.

In the first, we trained a **participant agnostic (PA) model** to predict participant choice without any identifying information about the participant or their behavior in other trials. The prediction problem is posed in the classical machine-learning format, to learn a predictive model: $M_{PA} = P(Y_t|X_t)$, where Y_t is the choice label for trial t , and X_t is the corresponding feature vector.

In the second paradigm, we trained an ensemble of **behavioral-typology (BT) models** that first group participants based on several rounds of trial data (including choice labels). The model is then required to predict participant choice for all remaining trials. This paradigm was motivated by the qualitative observation that participants' choices often became predictable to the experimenter once a strategic pattern was detected.

As such, the behavioral-typology model is given trial data and labels from the first 3 adversarial trials completed by each participant. The prediction problem for each trial t becomes $M_{BT} = P(Y_t|X_t, X_{0,1,2}, Y_{0,1,2})$. Here, Y_t and X_t are defined as before, and $X_{0,1,2}$ and $Y_{0,1,2}$ are the concatenated data and labels respectively from the first three trials.

To test this model, we chose a simple heuristic based on the most choice-informative feature according to exploratory analysis of the training set: the proportion of eye fixations on the chosen card. Participants were assigned a behavioral-typology based on this metric ($< 33\%$, $< 34 - 66\%$, and $> 67\%$). Independent classifiers were trained for each of the three typologies of training participants. At test time, the fixation heuristic was again used to assign a typology to each test participant (based on an estimate using only the first 3 adversarial trials), and inference was performed with the associated classifier.

2.6.1 Training Paradigm. Both prediction problems used the same dataset which was produced by randomly assigning 50% of participants into each set ($N_{train} = 28$ and $N_{test} = 27$), and then building a feature matrix (see following Section 2.6.2) where each row corresponded to a single non-practice trial (excluding those without a successful card selection) from a single participant, with *set* (train vs. test) determined by the participant's assignment. As such, all models were tested using trials from novel participants, with no opportunity for train-test leakage. The resultant matrix contained 1200 trials of which 52% were used for training. To compare predictive performance and the success of obfuscation, independent models were trained on pre- and post-adversarial trials separately.

We tried a number of off-the-shelf classification algorithms for this task, and report results for two that performed best overall: Scikit-learn's implementation of the Random Forest Decision Tree (RFDT) and Gradient Boosted Decision Tree (GBDT) classifiers [28]. Due to the limited sample size and lack of a separate evaluation set, only minimal hyper-parameter tuning was performed in order to avoid overfitting the test sample.

To assess each model's performance we report accuracy scores (percent of trials correctly predicted) across all test participants' trials.

2.6.2 Feature Engineering.

Fixations and Gaze. To conform with the training paradigm which called for the generation of a single binary prediction per trial, we employed a number of common techniques to extract lower-dimensional feature vectors from the high-dimensional raw gaze and fixation data collected during each trial's decision phase.

Fixations, which were tallied with a start and stop timestamp and target object, were used to produce features indicating the fixation count (F_i) and fixation fraction ($\frac{F_i}{\sum_j F_j}$) for each key object in the scene (e.g. left card, right card, holder, table). Since each table card was expected to be a particularly informative region of interest, we also computed the minimum, maximum, and mean duration of fixations on each. Additional fixation-based features included: last and second to last fixation object, and percent of trial for which eyes were closed.

In contrast to discrete fixations, features based on raw gaze points have the potential to capture behavior in which participants look near but avoid direct fixation upon an object of interest. We computed simple descriptive features from the mean, mode, and standard deviation of each of the three coordinate dimensions.

Head and Controller Motion. To capture dynamics of the controller and head mounted display (HMD), which measures head position and rotation, we included descriptive statistics from the position and rotation of each (e.g. HMD Y-position standard deviation, controller yaw mean, final HMD roll). Position and rotation features included an absolute version, as well as one relative to trial start. We estimated mean and maximum velocity based on position deltas and timestamps.

Finally, to capture more granular information that might indicate the trajectory of head and arm motion, we broke the decision phase into 10 equal duration segments, and included features for 3-axis position snapshots at each time point.

3 RESULTS

Our analysis indicates significant changes when comparing behavior between adversarial and non-adversarial trials, as well as substantial individual differences during performance of the task. In Section 3.2 we review results, including the post-task questionnaires and interviews, related to apparent behavioral strategies and avoidance of prediction.

3.1 Effect of Adversary

In the following sections we report results of analyses comparing data before and after each participant received the adversarial instruction. Unless stated, analyses include participants from both

early adversary and delayed adversary conditions. The physiological and behavioral variables analyzed exist on a vague continuum from signals typically considered involuntary (e.g. heart rate variability and electrodermal activity), to those usually interpreted as the result of intentional, conscious action (e.g. gross arm movements). The very premise of this study rests on the ambiguous middle ground, where participants may attempt to control or alter behaviors they know to be under observation in order to obfuscate their true intent. Which hand motions, head tilts, or saccades are the product of voluntary control versus involuntary physiological or motor outputs cannot be established by our analyses, or indeed any algorithm [16]. Though post-task interviews do shed light on the strategies consciously selected by participants, self-reports provide only a subset of all effects on behavior exhibited. For the purposes of this study, however, we need not discriminate voluntary versus involuntary action, and rather ask only how measurable behaviors change, and to what extent these changes affect the ability of an algorithmic agent to successfully predict a participant’s ultimate choice. For an analysis of behavior changes consistent with self-reported strategies, see Section 3.2.

3.1.1 Physiology. Though a slight peak is observed after adversarial trials start, we found no statistically significant difference in heart rate between non-adversarial and adversarial trials. Results suggest that the number of EDA spikes reduced post adversary ($t = 1.92, p = 0.059$), though temporal and learning effects are additional potentially confounding causes (see Figure 3a and 3b).

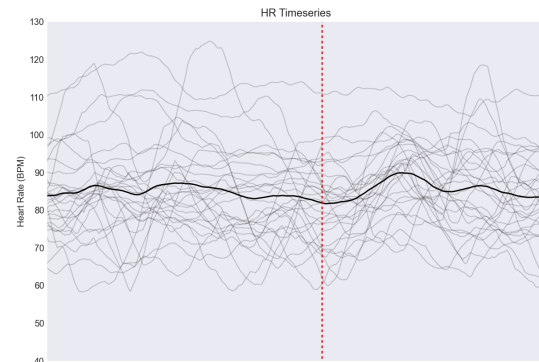
3.1.2 Movement. In adversarial trials, participants tended to move their arm and head less. Specifically, we assessed the range of motion along each rotational dimension (pitch, roll, yaw), as $\theta_{range} = \max(\theta) - \min(\theta)$. Range of all three dimensions for both controller (arm) and HMD (head) was significantly lower in post-adversary trials (statistical test results shown in Figure 4).

3.1.3 Gaze. The distribution of raw gaze data also significantly shifted during adversarial rounds. Both the mean and standard deviation of gaze target height ($y_{mean, t} : -3.55, p < 0.001$; $y_{SD, t} = -3.55, p < 0.001$) were significantly lower with adversary.

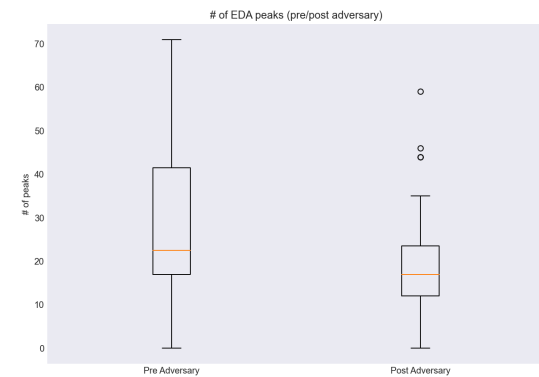
3.2 Strategic Behavior and Prediction

3.2.1 Self-Reported Strategies. Post-task questionnaires and interviews probed participants on how they interpreted the requirements of the post-adversary task, how they modified their behavior (if at all) to avoid prediction, as well as the experience of being asked to make a decision under adversarial surveillance. While the primary motivation for these interviews was to verify participants’ comprehension of the task, they contain some useful insights into conscious strategies used, as well as the challenges posed by the task overall.

Questionnaire responses indicate that participants primarily adjusted behavior related to gaze and arm motion (see Figure 5). In interviews, participants identify a range of meta-strategies which we categorize as signal dampening, signal randomizing and overt misdirection. Signal dampening strategies included the suppression of behaviors that participants believed may give away their intent, such as holding their head still, reducing motion in their



(a) Heart rate time-series for all delayed condition participants, across full session duration. Red dashed line indicates adversarial transition.



(b) Count of EDA spikes, pre- and post-adversary. Reduction in spikes was not statistically significant according to independent t -test.

Figure 3: Analysis of physiology data before and after adversary.

arm, pointing their hand in a neutral direction, entraining gaze on a neutral portion of the table or on the timer, and looking near but not directly at the cards on the table. Signal randomizing techniques included both adding excess behaviors during each trial, such as shifting gaze widely and continuously, moving their head back and forth at random, as well as randomizing overt asymmetrical behaviors by pointing the controller sometimes at the chosen card, and sometimes away.

Overt misdirection included intentional actions that ‘telegraph’ intent opposite to true choice, such as fixating gaze or head position for long durations on the opposite card, moving the controller towards the opposite card during the decision phase, etc.

Several participants reported initially looking at or near the table cards just long enough to make their decision, then switching to a misdirection or obfuscation strategy for the remainder of the

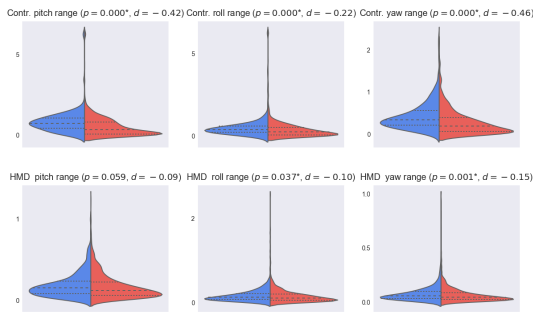


Figure 4: Head and controller range pre- (blue) and post-adversary (red). Range for both reduced significantly post-adversary suggesting support for signal dampening strategy. p -value for independent sample t -test, and Cohen’s d shown above each plot.

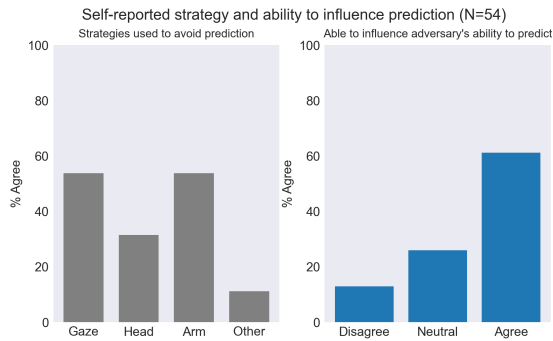


Figure 5: Results from post-task survey. Left: Participants reported modifying multiple behaviors with gaze and arm motion most often selected. Right: 61% of participants responded 'Agree' or 'Strongly Agree' when asked if they believed they were able to successfully influence the adversary’s ability to predict their choice.

decision phase: “I would look at both and figure out what the right answer was, and then I would stare at the wrong one, or go off into my own thoughts” (PP 012). Other participants talked about an evolving thought process on obfuscation, and the realization that consistent misdirection may also be predictable: “At first I thought, oh it makes sense to try not to do what you would normally do, like if you’re going to pick this, then look at the other one for longer. After a while I was like, oh that’s also predictable, so maybe switch it up.” (PP 015).

Over 60% of participants believed they were able to influence the adversary’s ability to predict their intent using the strategies they employed.

3.2.2 Agency and Cognitive Load. While the majority of participants reported that they believed they were able to influence the adversary’s prediction, some weren’t sure, and others disagreed.

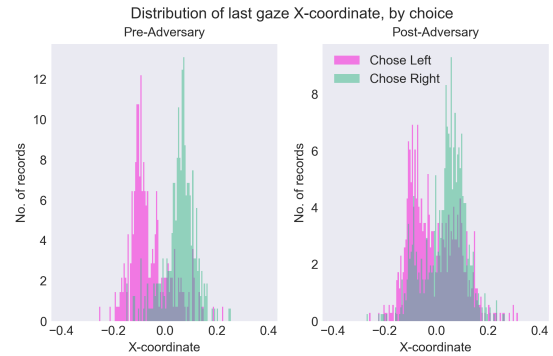


Figure 6: The increasing overlap of the gaze x-coordinate distribution between left vs. right trials post-adversary illustrates behavior change to deter prediction based on gaze. On average, however, final gaze point was highly predictive of choice both pre- and post-adversary.

In post-task interviews, some of these participants noted that they didn’t understand how the adversary was making its predictions, and so felt it wasn’t clear how they could effectively respond (PP 046, PP 016, PP 057): “I guess I just didn’t really know what it was going to be looking at, so I didn’t know what to change” (PP 017). Others found the matching aspect of the task challenging, and felt they needed to focus on selecting the right card rather than avoiding prediction: “I was more preoccupied with getting the right card, so kind of forgot about [the adversary] sometimes” (PP 003). For another, the decision and the mechanics of making their selection took priority: “I was still trying to pick the right card, like grab it in the right amount of time” (PP 016).

These assessments highlight that for some participants uncertainty about the dynamics of algorithmic prediction undermined their feeling of agency. Additionally, the cognitive load imposed by the combination of the card matching task and time limit were challenging for some participants to balance with the need to obfuscate. In the following section we review results comparing biometric data when left is chosen versus when right is chosen, thus identifying choice-correlated asymmetries likely exploited by the predictive model.

3.2.3 Correlates of Choice. To detect behavioral patterns that might be leveraged by a predictive model, we performed a variety of exploratory analyses comparing raw data and features collected during the decision phase, segmented by participant choice. Any asymmetries seen either across the full training set, or within a substantial sub-group, might indicate strategic behaviors or involuntary predictors correlated with intent. We report several such asymmetries below.

Gaze. Even when averaging across the full training population, the x-coordinate of the final raw gaze point was highly predictive of choice ($d = -0.447$, $t = -8.68$, $p < 0.0001$, see Figure 6).

Longitudinal Observations. Though comparing summary statistics of gaze and motion data highlights systematic differences in behavior, this approach ignores the richness of temporal visual

attention inherent to this decision-making paradigm. By analyzing individual participants movement and gaze longitudinally (both across trials, and within each trial), additional behavioral regularities appear.

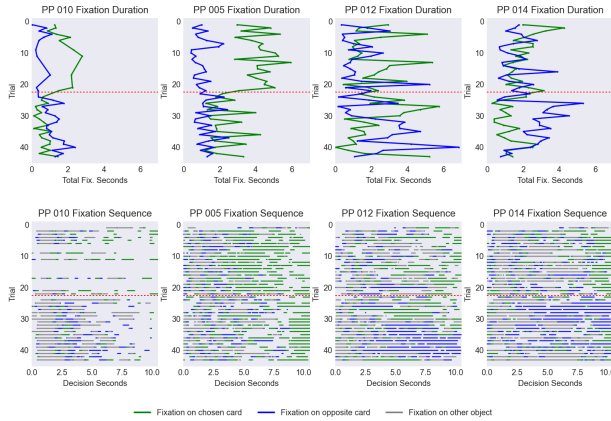


Figure 7: Total fixation duration (top) and fixation time-series (bottom) by trial for 4 selected participants (PPs). Green: fixations on chosen card; Blue: fixations on other card; Gray: fixations on non-card objects (shown in sequence charts only). Trials progress vertically from top to bottom, and the dotted red line indicates the beginning of adversary trials. Note that trials for which no card was selected are omitted.

Figure 7 illustrates trial-wise fixation time-series and total duration by card (chosen versus not chosen) for selected individual participants. These plots demonstrate dynamics of fixations, and behavioral change post-adversary, found to be representative of common behavioral patterns among the training population.

Participants PP 010 and PP 005 both favored fixations on the card they eventually chose prior to adversary, but show very different behavior post-adversary. Post-adversary, PP 005 reduces fixation duration on their chosen card, but both fixation duration, and final fixation (note the regularity in final fixation) still indicates choice on nearly every post-adversary trial. PP 010, on the other hand, is likely using a misdirection strategy and intentionally switches visual attention to the opposite card.

PP 014’s fixation behavior illustrate a key finding—in some cases, conscious obfuscation may unintentionally unmask intent. While fixations are well-balanced pre-adversary, the participant’s choice becomes significantly more predictable post-adversary under an apparent strategy of fixating on the opposite card.

Overall, these analyses highlight the value of providing temporal gaze sequence features (in addition to duration and fixation count) to the predictive models.

Arm Movement. An analysis of arm movement (measured by controller trajectories), reveals a substantial behavior change pre- and post-adversary. Trajectories are consistently stereotyped by choice in early non-adversarial trials, and show many participants preemptively (prior to selection phase) moving their controller over



Figure 8: Top-down view of controller trajectories (left), final position (middle) and final velocity (right), at termination of decision phase. Turquoise: right selected; Magenta: left selected. Clear choice correlation can be seen in all three metrics pre-adversary. Post-adversary data exhibits less stereotyped segmentation, but clear differentiation is still visible in both final position and final velocity.

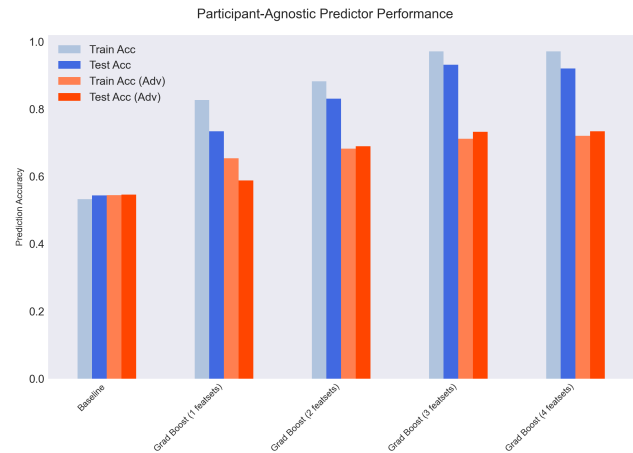


Figure 9: Model accuracy comparison for participant-agnostic predictor, with forward addition of feature-sets. Feature sets were added as follows: 1) Gaze features, 2) Fixation features, 3) HMD features, 4) Controller features.

the chosen card in preparation for selection. In post-adversary trials, we see support for self-reported strategies involving constrained controller motion as well as misdirection (e.g. the density of final controller positions over the opposite card in the bottom center plot in Figure 8).

Despite this visible obfuscation behavior, choice-correlation remains in final controller position, and particularly in final velocity. We discuss these findings further in Section 4.

3.2.4 Algorithmic Prediction.

Participant-Agnostic Predictor. Even in the more challenging participant-agnostic paradigm where the predictive model must

identify features successfully correlating with choice without the context of prior participant behavior, prediction rates were high. The best-performing model was the Gradient Boosted Decision Tree (GBDT), with a prediction accuracy of 93% on pre-adversary trials, and 73% on post-adversary trials (see Figure 9).

The most predictive features included last gaze fixation target, mean gaze X-coordinate, and final HMD roll offset.

Behavioral-Typology Predictor. As anticipated, the behavioral-typology (BT) predictor improved overall prediction accuracy, achieving 72.4% for the most effective (least predictable) strategy typologies, and 82.7% for the least effective behavioral typologies (see Figure 11 for BT model performance details).

The majority of participants are reliably predicted in between 75 to 100% of trials using this technique (see accuracy distribution post-adversary in Figure 10).

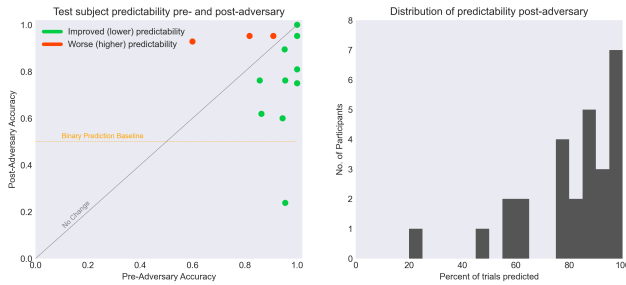


Figure 10: BT model predictability pre-adversary (x-axis) compared to post-adversary (y-axis). Some test participants (red) became more predictable post-adversary. The distribution of predictability (right) illustrates that the majority of participants’ choices could be successfully predicted in between 75% to 100% of their trials. Note: the chart at left includes only test participants in the delayed adversary condition, since BT model performance cannot be assessed from the three pre-adversary trials. The chart at right includes all test participants across both conditions.

Table 1 provides a summary of model prediction accuracy.

Strategic Efficacy. We can quantify the efficacy of a participant’s change in behavior by comparing our trained model’s prediction performance between pre- and post-adversary trials. We define a metric for strategic efficacy (Ψ) as follows:

$$\Psi = \frac{N_{correct,pre}}{N_{total,pre}} - \frac{N_{correct,post}}{N_{total,post}}$$

If efforts to conceal intent are effective, prediction accuracy post-adversary will be lower (positive Ψ). Figure 10 plots BT model accuracy pre- and post-adversary thus illustrating strategic efficacy (distance from diagonal). Participants for whom prediction accuracy decreased are plotted in green, while increases in prediction accuracy are plotted in red. The clustering of test participants below the diagonal illustrates that, through strategic adjustment to behavior, most participants were able to successfully reduce choice prediction, but not by much. Indeed, only one participant (shown

Table 1: Summary of predictive model performance. Best-performing models for each prediction paradigm are listed for both pre-adversary (pre) and post-adversary (post) trials, Gradient Boosted Decision Tree (GBDT) for the Participant-Agnostic (PA) predictor, and Random Forest Decision Tree (RFDT) for the Behavioral-Typology (BT) predictor.

Model	Train Accuracy	Test Accuracy
GBDT (PA, Pre)	97.1%	92.0%
GBDT (PA, Post)	72.0%	73.4%
RFDT (BT, Pre)	–	91.0%
RFDT (BT, Post)	–	80.4%

at lower right) evaded prediction in more than half of trials, and on average across participants, 80.4% of post-adversary trials were successfully predicted.

Also visible in Figure 10, is the surprising finding that for a number of participants (red dots), efforts to become less predictable had the opposite effect producing an increased prediction accuracy post-adversary.

To further investigate behavioral and physiological differences that might explain variance in strategic efficacy, we compared two groups of test participants: those less than 70% predictable post-adversary, which we term ‘effective’, and those at least 70% predictable post-adversary, which we term ‘ineffective’. The latter group can be seen as the rightmost disjoint cluster in the histogram in Figure 10. We find that the effective group exhibited statistically significant reduction in controller movement (yaw std. dev.: $d = 1.544, t = 3.012, p = 0.006$; final pitch: $d = 0.932, t = 9.950, p < .001$; roll range: $d = 0.583, t = 5.469, p < .001$), head movement (pitch std. dev.: $d = 0.745, t = 6.964, p < .001$; roll range: $d = 0.848, t = 8.111, p < .001$; yaw range: $d = 0.692, t = 6.758, p < .001$), percent of fixations on chosen card ($d = 1.470, t = 2.588, p = 0.016$) and total fixation duration on chosen card ($d = 1.263, t = 2.205, p = 0.037$). These results together indicate that, on average, the effective group moved their head and arm less during the decision phase, while also fixating less frequently and for shorter durations on the card they ultimately selected.

4 DISCUSSION

4.1 Behavior and Prediction

The experimental paradigm presented here was designed to provide insights into both the sensitivity of behavioral and biometric data when leveraged towards the prediction of future decisions (RQ1/2), the response of individuals confronted with a specific kind of adversarial surveillance (RQ3), as well as the interaction between the two in the effect of awareness of tracking on an individual’s ability to maintain private intentions (RQ4).

Classification accuracy results indicate that collected data were extremely predictive of choice (over 90% accuracy) for participants during pre-adversary trials. Though pre-adversary prediction was performed primarily to produce a baseline against which to compare performance during adversarial trials, this finding alone deserves

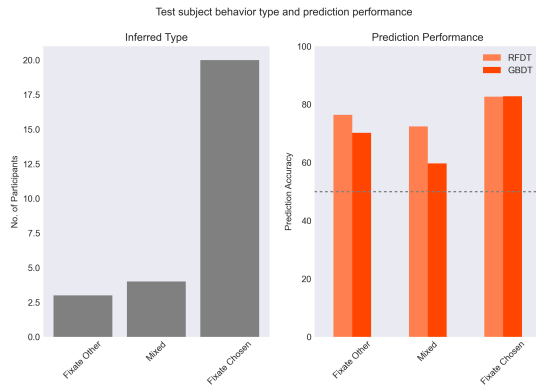


Figure 11: Behavioral-Typology model performance for each inferred typology based on proportion of fixations: fixations prefer chosen card, fixations prefer other card, and mixed.

some consideration. In the absence of any reason to mask intent, behavioral data collected during deliberation was sufficient to predict eventual choice in the vast majority of trials.

In adversarial trials, after being informed of the imperative to decide which card to select without revealing intent, most participants adopted strategies based on logical assumptions about the dynamics of prediction. Participant interviews revealed a thought process involving the various data they believed might be used by the adversary. The gaze calibration exercise at the beginning of the VR session suggested to many participants that their eye movements may be among the data used by the adversary to anticipate their decision. Seeing their controller in VR was another strong signal that arm motion might be a target of tracking, as well as an opportunity for masking or obfuscation. Some participants observed that during pre-adversary trials their hand would begin to move just before the end of the decision phase, and reported attempting to adjust, invert, or randomize this behavior post-adversary. Despite correctly identifying these signals, however, components of gaze, head, and arm motion all remained correlated with choice when averaged across the test sample. These correlations, in aggregate, allowed the classifiers within both prediction paradigms to successfully predict card choice in spite of the strategies employed.

To understand why, it is instructive to revisit one of the most informative data types captured: gaze fixations. In Figure 7, PP 010 and PP 005 both demonstrate significant adjustments to gaze dynamics upon the beginning of adversarial trials. Both strategies, once identified, remain predictive of choice—PP 005 reduces overall fixations on either card, and appears to randomize fixations while deciding, but fixates their chosen card at the end of each decision phase (green lines at end of post-adversary trials). PP 010 also reduces overall card fixations, and inverts their gaze behavior in a way that remains highly correlated with choice.

Optimal behavior, as any child fond of rock-paper-scissors (or game theory) intuitively knows, depends upon the randomization of behavior in order to minimize correlations between observable data and intended action. Randomization, even of a single series of discrete choices, however, is known to be impossible for humans

to achieve [32]. In contrast to univariate randomization, our task is especially challenging due to its multi-dimensionality—it requires successful randomization of all detectable behaviors simultaneously.

Finally, several of the strategies detected appear consistent with the types of intuitive behavior that might be effective at masking intent from a human observer. This is one interpretation of results showing lowered head pitch (Figure 4), and lowered gaze, which fail to obfuscate the target of gaze from eye-tracking hardware like that used in this study.

4.2 Limitations and Future Work

The prediction accuracy results achieved in our analysis are not meant to establish a ceiling. Rather than developing an optimal behavior prediction tool, the aim of this work was to assess the feasibility of intent prediction, as well as the range of individuals' responses to an explicitly surveilled task paradigm. With a larger training set, it is likely that even the standard statistical techniques used in this analysis would learn an improved behavioral model capable of exploiting the regularities of a wider variety of strategies and therefore achieve prediction accuracy improvements.

An additional limitation stems from the simplistic behavioral-typology inference model which we based on a single metric (card fixation ratio) indicative of a single behavioral strategy. With a slightly larger participant pool, an unsupervised learning approach identifying natural subject clusters or axes of behavioral variation would likely have produced improved prediction accuracy for the BT model.

While our results suggest dynamics that may extend into real-world contexts in which individuals interact with surveillance systems, the typical considerations of the generalizability of in-lab findings apply here. While a significant performance-based incentive was used to motivate task success, behavior in a setting with more significant potential consequences—as is increasingly relevant to real-world surveillance—might well display dynamics different from those observed in this study. Relatedly, though the game-based setting was conducive to our analysis, studying participants as they make more naturalistic decisions (e.g. who to vote for, or how much to tip) may offer insights that more easily extend to everyday behavior.

We hope future work will build on this preliminary study to develop a more holistic understanding of individuals' reasoning about and response to biometric surveillance.

4.3 Implications

The kinds of privacy risk supported by our findings are likely best addressed through a combination of public awareness, regulatory action, and the concerted efforts of designers [2, 29, 36].

Among other proposed solutions, Bailenson suggests users may take it upon themselves to use hardware filters capable of adding noise and reducing the fidelity of collected data [3]. Our findings, however, suggest that users may overestimate the efficacy of more intuitive obfuscation strategies, and underestimate the sensitivity of the data collected from them. As such, hoping to encourage costly user action such as obtaining and using obfuscation devices may not be a reliable solution.

Public education, however, must certainly play a role. Alphabet’s “Digital Transparency in the Public Realm” initiative is one example that hopes to encourage cities and other actors collecting data in public spaces to use a set of colored privacy icons to indicate the types and sensitivity of data being collected [14]. The behaviors and misconceptions illuminated by our results, however, indicate some critical limitations of this effort as well. What will an individual who notices an eye-tracking symbol on a light-post do with this information? Will Facebook require users of its recently announced AR glasses to wear warning labels to inform passersby of ongoing tracking [11]? What use is informing members of the public about surveillance, without instructions for opting-out? And if, as was the case for some of our participants, strategies to protect one’s privacy instead act to expose sensitive information, might messaging like this cause additional harms?

Even more fundamentally, however, any response to tracking technologies must contend with the fact that what is being collected is a moving target: the signals hidden within raw data evolve with new algorithms, new users, and integrations with additional data sources. The sensitivity emerges not from the data itself, but what can be done with it. This study highlights the counterintuitive nature of “what can be done,” and as such the challenges privacy advocates must overcome when communicating these ideas in the public sphere.

4.4 Ethical Considerations

It is crucial, when conducting work either employing or studying technologies that have been used to harm individuals and communities, to critically exam the potential impact on these populations and society as a whole. We reason that this research is unlikely to reinforce harms, and holds the potential to bring valuable data to academic and public dialogue regarding the capabilities, risks, and vulnerabilities that may be exploited by algorithmic surveillance.

First, this work does not develop any novel algorithm or computational model that might improve predictive performance, but rather explores the efficacy of off-the-shelf machine learning tools. Second, that the findings of this work likely parallel research conducted privately within organizations seeking to leverage algorithmic prediction in the interest of financial or political gain. If our speculation related to potential prediction accuracy is correct, we must acknowledge the multiple orders of magnitude that separate the sample size and scope of data used in this study with that available to companies and other organizations already in the business of collecting biometric data from individuals.

5 CONCLUSION

In this work, we examined the expectations individuals hold about biometric surveillance, and how these beliefs influence behavioral response to a tracked setting. Our results suggest that participants hold a range of priors about the nature of biosignals that might be leveraged for prediction, and use a wide variety of strategies to attempt to make a future choice less predictable. While some participants questioned their agency in evading the adversary, most modified their behavior and successfully reduced their prediction accuracy. However, data collected remained highly predictive of choice (over 80% mean accuracy), and the majority of participants

were correctly predicted by the behavioral-typology model in 75–100% of trials. Importantly, a meaningful subset of participants adopted a strategy that on average increased the model’s ability to successfully predict their choice, suggesting the counterintuitive nature of the dynamics of algorithmic prediction.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. 2019236659. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

- [1] Intiaz Ahmad, Rosta Farzan, Apu Kapadia, and Adam J Lee. 2020. Tangible privacy: Towards user-centric sensor designs for bystander privacy. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–28.
- [2] Claudio A Ardagna, Marco Cremonini, Sabrina De Capitani di Vimercati, and Pierangela Samarati. 2009. An obfuscation-based approach for protecting location privacy. *IEEE Transactions on Dependable and Secure Computing* 8, 1 (2009), 13–27.
- [3] Jeremy Bailenson. 2018. Protecting Nonverbal Data Tracked in Virtual Reality. *JAMA Pediatrics* 172, 10 (Oct. 2018), 905. <https://doi.org/10.1001/jamapediatrics.2018.1909>
- [4] Kirstie Ball. 2010. Workplace surveillance: An overview. *Labor History* 51, 1 (2010), 87–106.
- [5] Rhenan Bartels and Tiago Peçanha. 2020. HRV: a Pythonic package for Heart Rate Variability Analysis. *Journal of Open Source Software* 5, 51 (2020), 1867.
- [6] Sian L. Beilock and Lauren E. Holt. 2007. Embodied Preference Judgments: Can Likeability Be Driven by the Motor System? *Psychological Science* 18, 1 (Jan. 2007), 51–57. <https://doi.org/10.1111/j.1467-9280.2007.01848.x>
- [7] Clint Boulton. 2013. Snackmaker modernizes the impulse buy with sensors, analytics. *CIO Journal* 11 (2013).
- [8] Joseph Bullington. 2005. ‘Affective’ computing and emotion recognition systems: the future of biometric surveillance?. In *Proceedings of the 2nd annual conference on Information security curriculum development*. 95–99.
- [9] David Chaum. 2004. Secret-ballot receipts: True voter-verifiable elections. *IEEE security & privacy* 2, 1 (2004), 38–47.
- [10] Tamara Denning, Zakariya Dehlawi, and Tadayoshi Kohno. 2014. In situ with bystanders of augmented reality glasses: Perspectives on recording and privacy-mediating technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2377–2386.
- [11] Facebook. 2020. Announcing Project Aria: A Research Project on the Future of Wearable AR. <https://about.fb.com/news/2020/09/announcing-project-aria-a-research-project-on-the-future-of-wearable-ar/>.
- [12] Karl Friston, Rick A. Adams, Laurent Perrinet, and Michael Breakspear. 2012. Perceptions as Hypotheses: Saccades as Experiments. *Frontiers in Psychology* 3, MAY (2012), 1–20. <https://doi.org/10.3389/fpsyg.2012.00151> arXiv:NIHMS150003
- [13] Natalie Gagliardi. [n.d.]. Amazon Intros Alexa for Residential, Eyes Property Management Firms. <https://www.zdnet.com/article/amazon-intros-alexa-for-residential-eyes-property-management-firms/>.
- [14] Flavie Halais. [n.d.]. Is Your Data Being Collected? These Signs Will Tell You Where. *Wired* ([n. d.]).
- [15] Ian Hamilton. 2018. Privacy Summit At Stanford Will Draft VR ‘Bill Of Rights’.
- [16] Alicia Juarrero. 2000. Dynamics in action: Intentional behavior as a complex system. *Emergence* 2, 2 (2000), 24–57.
- [17] Malcolm Kirkup and Marylyn Carrigan. 2000. Video surveillance research in retailing: ethical issues. *International Journal of Retail & Distribution Management* (2000).
- [18] Nathan F. Lepora and Giovanni Pezzulo. 2015. Embodied Choice: How Action Influences Perceptual Decision Making. *PLoS Computational Biology* 11, 4 (April 2015), e1004110. <https://doi.org/10.1371/journal.pcbi.1004110>
- [19] Dongheng Li, David Winfield, and Derrick J Parkhurst. 2005. Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)-Workshops*. IEEE, IEEE, San Diego, California, USA, 79–79.
- [20] Daniel J. Liebling and Sören Preibusch. 2014. Privacy Considerations for a Pervasive Eye Tracking World. *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct Publication - UbiComp ’14 Adjunct* (2014), 1169–1177. <https://doi.org/10.1145/2638728.2641688>

- [21] Mirca Madianou. 2019. The biometric assemblage: Surveillance, experimentation, profit, and the measuring of refugee bodies. *Television & New Media* 20, 6 (2019), 581–599.
- [22] Avi Marciano. 2019. Reframing biometric surveillance: from a means of inspection to a form of control. *Ethics and Information Technology* 21, 2 (2019), 127–136.
- [23] Nick Merrill, John Chuang, and Coye Cheshire. 2019. Sensing is Believing: What People Think Biosensors Can Reveal About Thoughts and Feelings. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 413–420.
- [24] Betsy Morris. 2016. Virtual-Reality Startup Strivr Raises \$5 Million in Initial Funding Round. *Wall Street Journal* (Dec. 2016).
- [25] Lucas Hermann Negri and Christophe Vestri. 2017. *lucashn/peakutils: v1.1.0*. <https://doi.org/10.5281/zenodo.887917>
- [26] Helen Nissenbaum. 2009. *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.
- [27] Anjul Patney, Marco Salvi, JooHwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 179.
- [28] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [29] James Pierce. 2019. Smart home security cameras and shifting lines of creepiness: A design-led inquiry. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [30] Thammathip Piumsomboon, Gun Lee, Robert W Lindeman, and Mark Billinghurst. 2017. Exploring natural eye-gaze-based interaction for immersive virtual reality. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 36–39.
- [31] Inioluwa Deborah Raji, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. 2020. Saving face: Investigating the ethical concerns of facial recognition auditing. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 145–151.
- [32] Amnon Rapoport and David V Budescu. 1992. Generation of random series in two-person strictly competitive games. *Journal of Experimental Psychology: General* 121, 3 (1992), 352.
- [33] Franziska Roesner, Tadayoshi Kohno, and David Molnar. 2014. Security and privacy for augmented reality systems. *Commun. ACM* 57, 4 (2014), 88–96.
- [34] Elaine Sedenberg, Richmond Wong, and John Chuang. 2017. A window into the soul: Biosensing in public. (2017). [arXiv:1702.04235](https://arxiv.org/abs/1702.04235)
- [35] Anil K Seth. 2014. *The cybernetic Bayesian brain*. Open MIND. Frankfurt am Main: MIND Group.
- [36] Vincent Toubiana, Lakshminarayanan Subramanian, and Helen Nissenbaum. 2011. TrackMeNot: Enhancing the privacy of Web Search. *arXiv e-prints* (Sept. 2011). [https://doi.org/10.1016/S0370-2693\(96\)01648-6](https://doi.org/10.1016/S0370-2693(96)01648-6) [arXiv:1109.4677](https://arxiv.org/abs/1109.4677) [cs.CR]
- [37] Roel Vertegaal and Jeffrey S Shell. 2008. Attentive user interfaces: the surveillance and sousveillance of gaze-aware objects. *Social Science Information* 47, 3 (2008), 275–298.
- [38] Shoshana Zuboff. 2015. Big other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* 30, 1 (2015), 75–89.