# Technical Appendix for "AI-Assisted Scientific Data Collection with Iterative Human Feedback"

Travis Mandel, James Boyd, Sebastian J. Carter, Randall H. Tanaka, Taishi Nammoto

## 1 Introduction

This technical appendix contains three pieces:

1. A table of mathematical notation.
2. Full proofs of all results stated in the main paper.
3. Additional details on the user study.

We also repeat some information (e.g. problem setup) from the main paper for improved readability.

## 2 Table of notation

| Symbol | Meaning |
|---|---|
| $\mathcal{H}$ | Human/Scientist |
| $\mathcal{A}$ | AI agent |
| $\mathcal{E}$ | Environment |
| $N$ | The number of x-values |
| $\mathcal{X}$ | Set of all possible x-values $\{1, \ldots, N\}$ |
| $j_t$ | x-value chosen at timestep $t$ |
| $y_t$ | y-value received at timestep $t$ |
| $\mathcal{I}$ | The number of timesteps after which the human is periodically asked to provide feedback |
| $f_l(i)$ | lower bound function |
| $\hat{f}(i)$ | estimated function |
| $f_h(i)$ | upper bound function |
| $S_{i,t,C}$ | The set of y-values collected by algorithm $C$ associated with x-value $i$ at timestep $t$ (sometimes $t$ and/or $C$ are left off for convenience) |
| $\mathbf{S_{t,C}}$ | The vector of all $S_{i,t,C}$ at timestep $t$ for algorithm $C$ (sometimes $C$ is left off for convenience). |
| $\hat{K}_t$ | The latest set of keypoints $\mathcal{A}$ had received from $\mathcal{H}$ by timestep $t$. Keypoints are user-selected x-values. |
| $K$ | The true set of keypoints, which is unknown to both $\mathcal{H}$ and $\mathcal{A}$ initially |
| $n_i$ | The unknown maximum number of times we need to sample an x-value to determine whether it is a keypoint (see asssumption 1) |
| $\zeta(K, \mathbf{S_t})$ | The score function that measures how well we have sampled keypoints (want to maximize, pronounced "zeta") |
| $C(x)$ | A function returning the size of a confidence interval (takes in number of samples) |
| $\epsilon$ | The smoothing term used to prevent divide-by-zero in $C$ |
| $\delta$ | The confidence level used in $C$ |
| $d(\mathbf{S_t})$ | $\sum_{i \in K} C(\lvert S_{i,t} \rvert)$ |
| $z$ (Lemma 2) | A constant |
| $w$, $t_w$, $i_w$ (Lemma 3) | A y-value, the timestep it was collected, and its associated x-value |

| | |
|---|---|
| $q$, $t_q$, $i_q$ (Lemma 3) | A y-value, the timestep it was collected, and its associated x-value |
| $A$ | The theoretically optimal AI algorithm, see Theorem 1. |
| $H$ (Theorem 1) | a set containing all $i \in \mathcal{X}$ such that $\lvert S_{i,T,B}\rvert > \lvert S_{i,T,A}\rvert$ |
| $L$ (Theorem 1) | a set containing all $i \in \mathcal{X}$ such that $\lvert S_{i,T,B}\rvert \leq \lvert S_{i,T,A}\rvert$ |
| $t_i$ (Theorem 1) | For every $i$ in $H$, $t_i$ denotes the timestep such that $\lvert S_{i,t_i,B}\rvert = \lvert S_{i,T,A}\rvert$. For every $i'$ in $L$, $t_{i'}$ denotes the timestep where $\lvert S_{i',T,B}\rvert = \lvert S_{i',t_{i'},A}\rvert$ |
| $\eta_i$ (Theorem 1) | a set of all samples $y \in S_{i,T,B}$ where $y$ was collected at some $t' > t_i$ for $i \in H$ (for all $i \notin H$, let $\eta_i = \emptyset$) |
| $\eta$ (Theorem 1) | $\cup_{i \in H}\, \eta_i$ |
| $\Omega_{i'}$ (Theorem 1) | a set of all $y \in S_{i',T,A}$ where $y$ was collected at some $t' > t_{i'}$ for $i' \in L$ (for all $i \notin L$, let $\Omega_i = \emptyset$) |
| $\Omega$ (Theorem 1) | $\cup_{i \in L}\, \Omega_i$ |
| $m(y)$ Theorem 1) | A bijective function mapping from $\Omega$ to $\eta$ |
| $V'_C(y)$ (Theorem 1) | the increase in the function $d$ (see equation (3)) after collecting sample $y$ using an algorithm $C$ |
| $i_y$ | the x-value corresponding to the y-value $y$. |
| $t_y$ | the timesteps we received the y-value $y$. |
| $r'_D(t)$ | A typical notion of policy regret for algorithm $D$ at timestep $t$, defined in equation (50). We do not use this notion. |
| $A(\mathbf{S_{t-1,D}})$ | The sample algorithm $A$ would have received if it was initialized at $\mathbf{S_{t-1,D}}$ |
| $r_D(t)$ | Our notion of regret for algorithm $D$ at timestep $t$, defined in equation (52). Sometimes the subscript is left off for convenience. |
| $\epsilon_g$ (Corollary 1) | The probability the algorithm proposed in Corollary 1 selects a non-keypoint. |
| $\ell$ (Theorem 3) | A number of samples used in the statement of Theorem 3. |
| $\delta'$ (Theorem 3) | A fixed probability greater than 0. |
| $K_n$ (Theorem 3) | A subset of $K$ such that for all $i \in K_n$, $P(\lvert S_{i,T,D}\rvert \geq \ell') < 1 - \delta'$ as $T \to \infty$ for some $\ell'$ |
| $K_p$ (Theorem 3) | $K - K_n$ |
| $t'$ (Theorem 3) | the first timestep such that $\lvert S_{i,t',D}\rvert > \ell'$ for all $i \in K_p$ |
| $\tau$ (Theorem 3) | the set containing all $T'$ such that the chosen x-value $j_{T',D} \in K_p$ and $T' > t'$ |
| $\zeta_{min}$ (Theorem 3) | A constant defined in equation (79). |
| $i_c$ (Corollary 2) | The first keypoint. |
| $\ell_c$ (Corollary 2) | The number of samples of x-value $i_c$ after which it is revealed as a keypoint by $\mathcal{H}$. |
| $n_{max}$ | $\max_{i \in K} n_i$ |
| $\epsilon_p$ | TESA parameter specifying non-keypoint sample probability. |
| $b_0$ | TESA parameters specifying prior Pareto scale. |
| $a$ | TESA parameter specifying prior Pareto shape. |
| $\hat{K}$ | TESA variable keeping track of the most recent keypoints. |
| $b_i$ | TESA variable storing estimate of when each keypoint would have been added. |
| $x_m$ | TESA variable storing the posterior Pareto scale. |
| $h$ | TESA variable storing the sample from the Pareto distribution (estimate of $n_{max}$). |
| $i_k$ | TESA variable storing the least-sampled current keypoint. |
| $i_h$ | TESA variable storing the least-sampled x value. |
| $r$ | TESA variable storing a random floating point number. |
| $i_c$ | TESA variable holding the current x-value choice. |
| $X_p$ (Theorem 4) | the set of all $i \in \mathcal{X}$ such that $\lvert S_{i,t}\rvert \to \infty$ as $t \to \infty$. |
| $X_n$ (Theorem 4) | $\mathcal{X} - X_p$. |
| $\delta'$ (Theorem 4) | fixed probability greater than 0. |
| $\ell'$ (Theorem 4) | fixed bound on number of samples. |
| $t''$ (Theorem 4) | sufficiently large timestep such that $\lvert S_{i,t'',C}\rvert \leq \ell'$ for all $i \in X_n$ but, for all $i \in X_p$, $\lvert S_{i,t'',C}\rvert > \ell'$. |
| $\delta''$ (Theorem 4) | arbitrary chosen probability. |
| $t'$ (Theorem 4) | a sufficiently large timestep such that $\lvert S_{i,t'-1}\rvert > n_i$ for all $i \in \mathcal{X}$ with probability $1 - \delta''$ |

# 3 Problem Setup

We consider problems in which a human $\mathcal{H}$ and an AI $\mathcal{A}$ work together to gather the most scientifically useful data from an environment $\mathcal{E}$ (which maps x values to (noisy) y values). At each time step $t \geq 0$, $\mathcal{A}$ selects an x value $j_t \in \mathcal{X}$, where $\mathcal{X}$ is represented as the set of integer values $\{1, \ldots, N\}$, receives a sampled y value $y_t \in [0, 1]$ from $\mathcal{E}$, and adds it to a set of samples $S_{j_t}$. $\mathcal{H}$ is asked to provide feedback every $\mathcal{I}$ time steps. Specifically, when $t \mod \mathcal{I} = 0$, $\mathcal{A}$ will generate a visualization tuple $(f_h, f_l, \hat{f})$, where $f_h : \mathcal{X} \to [0, 1]$ is a higher bound function, $f_l : \mathcal{X} \to [0, 1]$ is a lower bound function and $\hat{f} : \mathcal{X} \to [0, 1]$ is an estimated function, which are estimated using the sample sets $S_i$ (for all $i \in \mathcal{X}$).[1] After observing the visualized tuple $(f_h, f_l, \hat{f})$, $\mathcal{H}$ sends $\mathcal{A}$ a set of keypoints $\hat{K}_t \subseteq \mathcal{X}$ (these are intended to represent scientifically interesting x-values of the curve, and the set will ideally grow as more information is uncovered about the true curve).

Let $K$ be the true set of keypoints, which is unknown to both $\mathcal{H}$ and $\mathcal{A}$ initially. We make the following assumption:

**Assumption 1.** *There exists some unknown $n_i$ such that if $|S_{i,t}| \geq n_i$ then $i \in \hat{K}_t$ if and only if $i \in K$.*

Note that Assumption 1 is not particularly strong, as it only requires $\mathcal{H}$ to make correct decisions in a long-term sense (i.e. if $|S_{i,t}| \geq n_i$). With smaller numbers of samples, $\mathcal{H}$ could make mistakes, for instance marking something a keypoint when it is not or deleting a true keypoint.

The goal of $\mathcal{A}$ is to optimize the following score (zeta):

$$\zeta\left(K, \mathbf{S_t}\right) = \frac{|K|}{\sum_{i \in K} C\left(|S_{i,t}|\right)} - \sum_{i \notin K} |S_{i,t}| \tag{1}$$

where $\mathbf{S_t}$ is the vector of sample set sizes for all x values at timestep $t$: $\langle |S_{1,t}|, \ldots, |S_{N,t}| \rangle$, and

$$C\left(x\right) = \sqrt{\frac{ln\left(\frac{2}{\delta}\right)}{2x + \epsilon}} \tag{2}$$

represents the confidence interval, $\epsilon > 0$ represents a (generally small) smoothing term, and $1 - \delta$ represents the confidence level given $\delta \in (0, 1]$. This is similar to the Chernoff-Hoeffding bounds, except for the addition of $\epsilon$ in the denominator to prevent divide-by-zero issues.[2]

The motivation for equation (1) is that we want more identified keypoints (hence the $|K|$ numerator), but we also want to be certain of the value of each keypoint (hence the sum of keypoint confidence interval sizes in the denominator). Also, we want to minimize the number of samples devoted to places that are not scientifically interesting, hence the term which subtracts the number of non-keypoint samples.

# 4 Defining Regret

Since our problem involves running AI systems online in an unknown environment, theoretical guarantees (such as showing that an algorithm is zero-regret) are particularly important. **Here we provide full proofs of the results stated in the main text.**

## 4.1 Properties of the score function $\zeta$

Here we show some properties of the score function $\zeta$ that will be useful for later results. First, we define a function $d$:

$$d\left(\mathbf{S_t}\right) = \sum_{i \in K} C\left(|S_{i,t}|\right) \tag{3}$$

In other words, $d$ is the denominator of the first term of $\zeta$.

**Lemma 1.** *$d$ is monotonically decreasing with $t$, and $\zeta(K, \mathbf{S_t})$ is monotonically increasing on $t$ such that $j_t \in K$.*

[1] In our experiments $\hat{f}(i) = \frac{\sum_{y \in S_i} y}{|S_i|}$, $f_l(i) = \hat{f}(i) - C(|S_i|)$, and $f_h(i) = \hat{f}(i) + C(|S_i|)$, where $C$ is as defined in equation (2).

[2] We let $\delta = 0.1$ and $\epsilon = 0.02$ in our experiments. Also, if $K = \emptyset$, we simply let $\zeta\left(K, \mathbf{S_t}\right) = 0$ to avoid divide-by-zero issues.

*Proof.* At timestep $t$, $y_t$ is added to $S_{j_t,t}$. Since $j_t \in K$, this must change the output of $d$. Analyzing the partial derivative of $d$ with respect to $|S_{j_t,t}|$ gives us:

$$\frac{\partial d}{\partial |S_{j_t,t}|} = -\sqrt{\ln\left(\frac{2}{\delta}\right)}\left(\frac{1}{2|S_{j_t,t}|+\epsilon}\right)^{1.5} \tag{4}$$

which means that $d$ monotonically decreases if an element is added to $S_{j_t,t}$. Now, $\zeta\left(K, \mathbf{S_t}\right) = \frac{|K|}{d(\mathbf{S_t})} - \sum_{i \notin K} |S_{i,t}|$. The last term remains constant on $t$ such that $j_t \in K$. So if $d$ monotonically decreases as samples are added, then $\zeta$ must monotonically increase on $t$ such that $j_t \in K$. □

Additionally, we show that the confidence interval function is decreasing and convex, that is, it shrinks less the more samples are gathered:

**Lemma 2.** *$C(x)$ is monotonically decreasing and convex for all $x \geq 0$.*

*Proof.* Let $z = \sqrt{\ln\left(\frac{2}{\delta}\right)}$. Note that since $0 < \delta \leq 1$, $z > 0$. Consider the first and second derivatives of the confidence interval function $C(x)$:
$\frac{dC}{dx} = -z\left(2x+\epsilon\right)^{-1.5}$ so $C$ is monotonically decreasing.
$\frac{d^2C}{dx^2} = 3z\left(2x+\epsilon\right)^{-2.5}$ so $C$ is convex. □

In the next lemma we show that there are diminishing returns, that is, if we choose a keypoint, the denominator shrinks less if there are more samples:

**Lemma 3.** *A sample $w \in [0,1]$, was collected at timestep $t_w$ at x-value $i_w \in K$, and similarly, a sample $q \in [0,1]$, was collected at timestep $t_q$ at x-value $i_q \in K$. Then for any two (possibly equal) algorithms $A$ and $B$ it must be that, if $|S_{i_w,t_w,A}| \leq |S_{i_q,t_q,B}|$, then*

$$d(\mathbf{S_{t_w,A}}) - d(\mathbf{S_{t_w-1,A}}) \leq d(\mathbf{S_{t_q,B}}) - d(\mathbf{S_{t_q-1,B}})$$

*Proof.* By Lemma 2 $C$ is convex and decreasing. If a function $f(x)$ is convex and decreasing and $x_0 < x_1$ then it must be that $f\left(x_0+1\right) - f\left(x_0\right) \leq f\left(x_1+1\right) - f\left(x_1\right)$, so:

$$C\left(|S_{i_w,t_w,A}|\right) - C\left(|S_{i_w,t_w,A}|-1\right) \leq C\left(|S_{i_q,t_q,B}|\right) - C\left(|S_{i_q,t_q,B}|-1\right) \tag{5}$$
$$C\left(|S_{i_w,t_w,A}|\right) - C\left(|S_{i_w,t_w-1,A}|\right) \leq C\left(|S_{i_q,t_q,B}|\right) - C\left(|S_{i_q,t_q-1,B}|\right) \tag{6}$$

$$\begin{aligned} &C\left(|S_{i_w,t_w,A}|\right) - C\left(|S_{i_w,t_w-1,A}|\right) + \left(\sum_{j\in K, j\neq i_w} C\left(|S_{j,t_w,A}|\right) - \sum_{j\in K, j\neq i_w} C\left(|S_{j,t_w,A}|\right)\right) \\ &\leq C\left(|S_{i_q,t_q,B}|\right) - C\left(|S_{i_q,t_q-1,B}|\right) + \left(\sum_{j\in K, j\neq i_q} C\left(|S_{j,t_q-1,B}|\right) - \sum_{j\in K, j\neq i_q} C\left(|S_{j,t_q-1,B}|\right)\right) \end{aligned} \tag{7}$$

Examining the left hand side of equation (7) gives us:

$$\begin{aligned} C\left(|S_{i_w,t_w,A}|\right) + \sum_{j\in K, j\neq i_w} C\left(|S_{j,t_w,A}|\right) - C\left(|S_{i_w,t_w-1,A}|\right) - \sum_{j\in K, j\neq i_w} C\left(|S_{j,t_w,A}|\right) \\ = \sum_{i\in K} C\left(|S_{i,t_w,A}|\right) - \sum_{i\in K} C\left(|S_{i,t_w-1,A}|\right) \end{aligned} \tag{8}$$

Since the only sample set that changed size from $t_w - 1$ to $t_w$ in algorithm $A$ was $S_{i_w,t_w,A}$. Similarly, examining the right hand side of equation (7) gives us:

$$C\left(|S_{i_q,t_q,B}|\right) + \sum_{j\in K, j\neq i_q} C\left(|S_{j,t_q-1,B}|\right) - C\left(|i_q, S_{t_q-1,B}|\right) - \sum_{j\in K, j\neq i_q} C\left(|S_{j,t_q-1,B}|\right) = \sum_{i\in K} C\left(|S_{i,t_q,B}|\right) - \sum_{i\in K} C\left(|i, S_{t_q-1,B}|\right) \quad (9)$$

Combining equations (7), (8) and (9) gives us:

$$\sum_{i\in K} C\left(|S_{i,t_w,A}|\right) - \sum_{i\in K} C\left(|S_{i,t_w-1,A}|\right) \leq \sum_{i\in K} C\left(|S_{i,t_q,B}|\right) - \sum_{i\in K} C\left(|S_{i,t_q-1,B}|\right) \quad (10)$$

By the definition of $d$:

$$d\left(\mathbf{S_{t_w,A}}\right) - d\left(\mathbf{S_{t_w-1,A}}\right) \leq d\left(\mathbf{S_{t_q,B}}\right) - d\left(\mathbf{S_{t_q-1,B}}\right) \quad (11)$$

□

## 4.2 Comparing Regret Definitions

Analyzing performance based on $\zeta$ itself is challenging, since the maximum achievable score depends on $\mathcal{E}$ and $\mathcal{H}$. Therefore, we analyze regret compared to an optimal algorithm. In this setting the optimal algorithm is not immediately obvious, so we introduce Theorem 1.

**Theorem 1.** *Let $A$ be an algorithm which selects the x value $\operatorname{argmin}_{i\in K} |S_i|$. $A$ is optimal according to the $\zeta$ score.*

*Proof.* If $|K| = 0$ all algorithms have a $\zeta$ score of 0, so here we consider cases where $|K| > 0$.
For a contradiction assume there is some algorithm $B$ and timestep $T$ such that $\zeta\left(K, \mathbf{S_{T,B}}\right) > \zeta\left(K, \mathbf{S_{T,A}}\right)$[3]
Let $H \subseteq \mathcal{X}$ be a set containing all $i \in \mathcal{X}$ such that $|S_{i,T,B}| > |S_{i,T,A}|$.
Let $L \subseteq \mathcal{X}$ be a set containing all $i \in \mathcal{X}$ such that $|S_{i,T,B}| \leq |S_{i,T,A}|$.
For every x-value $i$ in $H$ there exists some $t_i < T$ such that

$$|S_{i,t_i,B}| = |S_{i,T,A}| \quad (12)$$

Let $\eta_i$ be a set of all samples $y \in S_{i,T,B}$ where $y$ was collected at some $t' > t_i$ for some $i \in H$ (for all $i \notin H$, let $\eta_i = \emptyset$). Then we can express the final set $S_{i,T,B}$ in terms of $\eta_i$ and the earlier set $S_{i,t_i,B}$ as follows:

$$S_{i,t_i,B} \cup \eta_i = S_{i,T,B} \quad (13)$$

So we can write the cardinality as follows:

$$|S_{i,t_i,B}| + |\eta_i| = |S_{i,T,B}| \quad (14)$$

Using equation (12),

$$|S_{i,T,A}| + |\eta_i| = |S_{i,T,B}| \quad (15)$$

$$|\eta_i| = |S_{i,T,B}| - |S_{i,T,A}| \quad (16)$$

Let $\eta = \cup_{i\in H} \eta_i$. Then $|\eta| = \sum_{i\in H} |\eta_i|$, so:

$$|\eta| = \sum_{i\in H} |S_{i,T,B}| - |S_{i,T,A}| \quad (17)$$

[3]Note that in reality B might be stochastic, either directly or in the sense that the values of the samples might be stochastic. However, we here show that no algorithm will **ever** be able to allocate samples in a way that achieves better $\zeta$ score than $A$, which clearly implies that a stochastic algorithm cannot achieve better expected $\zeta$ score than $A$.

Likewise,for every x-value $i'$ in $L$ there exists some $t_{i'}$ where $|S_{i',T,B}| = |S_{i',t_{i'},A}|$. Letting $\Omega_{i'}$ be a set of all $y \in S_{i',T,A}$ where $y$ was collected at some $t' > t_{i'}$ for some $i' in L$ (for all $i' \notin L$, let $\Omega_i = \emptyset$). Using a similar argument to equations (12) - (16) above gives us:

$$|\Omega_{i'}| = |S_{i',T,A}| - |S_{i',T,B}| \tag{18}$$

Let $\Omega = \cup_{i \in L} \Omega_i$. Then $|\Omega| = \sum_{i \in L} |\Omega_i|$, so:

$$|\Omega| = \sum_{i \in L} |S_{i,T,A}| - |S_{i,T,B}| \tag{19}$$

Since both algorithms $A$ and $B$ have sampled $T$ times by timestep $T$:

$$\sum_{i \in \mathcal{X}} |S_{i,T,A}| = \sum_{i \in \mathcal{X}} |S_{i,T,B}| = T \tag{20}$$

Now, since $L$ and $H$ partition $\mathcal{X}$, it must be that

$$\sum_{i \in \mathcal{X}} |S_{i,T,A}| = \sum_{i \in L} |S_{i,T,A}| + \sum_{i \in H} |S_{i,T,A}| \tag{21}$$

and

$$\sum_{i \in \mathcal{X}} |S_{i,T,B}| = \sum_{i \in L} |S_{i,T,B}| + \sum_{i \in H} |S_{i,T,B}| \tag{22}$$

Combining equations (20), (21), and (22) gives us:

$$\sum_{i \in L} |S_{i,T,A}| + \sum_{i \in H} |S_{i,T,A}| = \sum_{i \in L} |S_{i,T,B}| + \sum_{i \in H} |S_{i,T,B}| = T \tag{23}$$

$$\sum_{i \in L} |S_{i,T,B}| + \sum_{i \in H} |S_{i,T,B}| - \sum_{i \in L} |S_{i,T,A}| - \sum_{i \in H} |S_{i,T,A}| = 0 \tag{24}$$

$$\sum_{i \in L} |S_{i,T,B}| - |S_{i,T,A}| + \sum_{i \in H} |S_{i,T,B}| - |S_{i,T,A}| = 0 \tag{25}$$

$$\sum_{i \in H} |S_{i,T,B}| - |S_{i,T,A}| = \sum_{i \in L} |S_{i,T,A}| - |S_{i,T,B}| \tag{26}$$

Combining equation (26) with equation (17) and equation (19) we get:

$$|\eta| = |\Omega|$$

Since $\eta$ and $\Omega$ are the same size, there exist many possible bijective functions mapping from $\Omega$ to $\eta$. Consider one such bijective function $m : \Omega \to \eta$.

Let the function $V'_C$ be defined as the increase in the function $d$ (see equation (3)) after collecting sample $y$ using an algorithm $C$, in other words define $V'_C$ as follows:

$$V'_C(y) = d\left(\mathbf{S_{t_y,C}}\right) - d\left(\mathbf{S_{t_y-1,C}}\right) \tag{27}$$

where $t_y$ denotes the timestep just after $y$ was collected (in other words, $S_{t_y,C}$ is the set which includes $y$ as a sample, while $S_{t_y-1,C}$ does not).

Our next step is to show $V'_B(m(y)) \geq V'_A(y)$ for all $y \in \Omega$. To do so, consider 2 cases:

Case 1: $i_{m(y')} \notin K$ Case 2: $i_{m(y')} \in K$

where $i_{m(y')} \in \mathcal{X}$ denote the x-value corresponding to the sample $m(y')$.

*Case 1:* $i_{m(y')} \notin K$

Since no samples were collected for any $i \in K$ at timestep $t$, $S_{i,t_{y'},B} = S_{i,t_{y'}-1,B}$ for all $i \in K$, and thus

$$d\left(\mathbf{S}_{\mathbf{t}'_{\mathbf{y}},\mathbf{B}}\right) = d\left(\mathbf{S}_{\mathbf{t}_{\mathbf{y}'-\mathbf{1}},\mathbf{B}}\right) \tag{28}$$

so $V'_B\left(m\left(y'\right)\right) = 0$.

Now, A always picks some $i \in K$, so $i_{y'} \in K$ and thus $d$ will change. By Lemma 1, $d$ monotonically decreases as additional samples are collected, and thus $V'_A\left(y'\right) \leq 0$. Since $V'_B\left(m\left(y'\right)\right) = 0$:

$$V'_A\left(y'\right) \leq V'_B\left(m\left(y'\right)\right) \tag{29}$$

*Case 2:* $m\left(y'\right) \in K$

Since $A$ selects $\operatorname{argmin}_{i \in K} |S_i|$, $A$ will never add more than 1 more sample to any x-value $j$ compared to any other x-value $j'$. So it must be that:

$$\left\lfloor \frac{T}{|K|} \right\rfloor \leq |S_{j,T,A}| \leq \left\lceil \frac{T}{|K|} \right\rceil \text{for all } j \in \mathcal{X} \tag{30}$$

$$|S_{i_{y'},T,A}| \leq \left\lceil \frac{T}{|K|} \right\rceil \tag{31}$$

$$|S_{i_{y'},t_{y'},A}| \leq \left\lceil \frac{T}{|K|} \right\rceil \tag{32}$$

Now, since $m\left(y'\right) \in \eta$, it must be that $i_{m(y')} \in H$ by definition of $\eta$. So $|S_{i_{m(y')},T,A}| < |S_{i_{m(y')},t_{m(y')},B}|$ by definition of $H$. Combining this with equation (30), we have:

$$\left\lfloor \frac{T}{|K|} \right\rfloor < |S_{i_{m(y')},t_{m(y')},B}| \tag{33}$$

$$\left\lceil \frac{T}{|K|} \right\rceil \leq |S_{i_{m(y')},t_{m(y')},B}| \tag{34}$$

Combining this with equation (32) gives us:

$$|S_{i_{y'},t_{y'},A}| \leq \left\lceil \frac{T}{|K|} \right\rceil \leq |S_{i_{m(y')},t_{m(y')},B}| \tag{35}$$

$$|S_{i_{y'},t_{y'},A}| \leq |S_{i_{m(y')},t_{m(y')},B}| \tag{36}$$

Using Equation (36), we can invoke Lemma 3, which gives us:

$$d\left(\mathbf{S}_{\mathbf{t}_{\mathbf{y}'},\mathbf{A}}\right) - d\left(\mathbf{S}_{\mathbf{t}_{\mathbf{y}'-\mathbf{1}},\mathbf{A}}\right) \leq d\left(\mathbf{S}_{\mathbf{t}_{\mathbf{m}(\mathbf{y}')},\mathbf{B}}\right) - d\left(\mathbf{S}_{\mathbf{t}_{\mathbf{m}(\mathbf{y}')-\mathbf{1}},\mathbf{B}}\right) \tag{37}$$

By the definition of $V'$:

$$V'_A\left(y'\right) \leq V'_B\left(m\left(y'\right)\right) \tag{38}$$

So by equations (29) and (38), $V'_A\left(y'\right) \leq V'_B\left(m\left(y'\right)\right)$ for all $y' \in \Omega$ regardless of which case we are in. Since $m$ is a bijection, this gives us that:

$$\sum_{o' \in \Omega} V'_A\left(o'\right) \leq \sum_{j' \in \eta} V'_B\left(j'\right) \tag{39}$$

Now, recall $\Omega_i$ contains all samples collected for x-value $i$ by $A$ after exceeding the corresponding number of samples for $i$ in $B$, and $\eta_i$ contains all samples collected for x-value $i$ by $B$ after exceeding the number of samples for $i$ in $A$. So for each $i$, it must be the case that

$$|S_{i,T,A} - \Omega_i| = |S_{i,T,B} - \eta_i| \text{ for all } i \in \mathcal{X} \tag{40}$$

Given equation (40) and the fact that $V'$ depends only on the number of samples at each x-value at the time a sample was collected, it must be that that samples aside from those in $\Omega$ and $\eta$ have equal total $V'$. In other words:

$$\sum_{o' \in \mathbf{S_{A,T}} - \Omega} V'_A(o') = \sum_{j' \in \mathbf{S_{B,T}} - \eta} V'_B(j') \tag{41}$$

Combining equations (39) and (41) gives us:

$$\sum_{y' \in \mathbf{S_{A,T}} - \Omega} V'_A(y') + \sum_{o' \in \Omega} V'_A(o') \leq \sum_{j' \in \mathbf{S_{B,T}} - \eta} V'_B(j') + \sum_{j' \in \eta} V'_B(j') \tag{42}$$

$$\sum_{o' \in \mathbf{S_{A,T}}} V'_A(o') \leq \sum_{j' \in \mathbf{S_{B,T}}} V'_B(j') \tag{43}$$

Recall from equation (27) that $V'_A(y) = d\left(\mathbf{S_{t_y}}, \mathbf{A}\right) - d\left(\mathbf{S_{t_y - 1}}, \mathbf{A}\right)$. By the fact that $\sum_{o' \in \mathbf{S_{A,T}}} V'_A(o')$ is a telescoping sum, we have that

$$\sum_{o' \in \mathbf{S_{A,T}}} V'(o') = d(\mathbf{S_{T,A}}) - d(\mathbf{I}), \text{ where } \mathbf{I} \text{ is the initial set vector } \langle \emptyset, \ldots, \emptyset \rangle \tag{44}$$

and likewise for $B$. Combining this with equation (43) we have

$$d(\mathbf{S_{T,A}}) - d(\mathbf{I}) \leq d(\mathbf{S_{T,B}}) - d(\mathbf{I}) \tag{45}$$

$$d(\mathbf{S_{T,A}}) \leq d(\mathbf{S_{T,B}}) \tag{46}$$

$$\frac{1}{d(\mathbf{S_{T,A}})} \geq \frac{1}{d(\mathbf{S_{T,B}})} \tag{47}$$

$$\frac{|K|}{d(\mathbf{S_{T,A}})} \geq \frac{|K|}{d(\mathbf{S_{T,B}})} - \sum_{i \notin K} |S_{i,T,B}| \tag{48}$$

By definition of $\zeta$, and the fact that $\sum_{i \notin K} |S_{i,T,A}| = 0$:

$$\zeta(K, \mathbf{S_{T,A}}) \geq \zeta(K, \mathbf{S_{T,B}}) \tag{49}$$

Therefore $B$ cannot be a better algorithm than $A$ according to the $\zeta$-score.

□

# 5 Defining Regret

Given Theorem 1, we wish to define regret of some algorithm $D$ relative to $A$. Perhaps the most natural definition of regret would be $\zeta(K, \mathbf{S_{T,A}}) - \zeta(K, \mathbf{S_{T,D}})$. However, since the $\zeta$ metric incorporates information from all timesteps (and not just the current one), this does not quantify how much regret we incur on each timestep $t$. This can be achieved by computing the change in score from $t-1$ to $t$ for both algorithms:

$$r'_D(t) = [\zeta(K, \mathbf{S_{t,A}}) - \zeta(K, \mathbf{S_{t-1,A}})] - [\zeta(K, \mathbf{S_{t,D}}) - \zeta(K, \mathbf{S_{t-1,D}})] \tag{50}$$

Note that, since our problem setup allows some quantities to be chosen adversarially (for instance, the addition and removal of keypoints), equation (50) measures regret with respect to the best possible policy in the face of these adversarial decisions, in other words, it is very similar to policy regret [1]. An alternative is to use a notion closer the external regret [2]:

$$r_D(t) = [\zeta(K, \mathbf{S_{t-1,D}} \cup A(\mathbf{S_{t-1,D}})) - \zeta(K, \mathbf{S_{t-1,D}})] - [\zeta(K, \mathbf{S_{t,D}}) - \zeta(K, \mathbf{S_{t-1,D}})] \tag{51}$$

$$r_D(t) = [\zeta(K, \mathbf{S_{t-1,D}} \cup A(\mathbf{S_{t-1,D}})) - \zeta(K, \mathbf{S_{t,D}})] \tag{52}$$

where $A(\mathbf{S_{t-1,D}})$ denotes the sample algorithm $A$ would have received if it was initialized at $\mathbf{S_{t-1,D}}$.[4]

In other settings, policy regret (i.e. Equation (50)) is thought to be strictly stronger than external regret (i.e. Equation (52)) [1]. Therefore, in most adversarial settings, regret guarantees with respect to the policy regret would be

[4] $\zeta$ does not depend on y-values, so one can look at the x-value A would select given $\mathbf{S_{t-1,D}}$ and increment the size of that set.

preferred to guarantees related to the external regret. But in this situation, policy regret is actually weaker in some sense than external regret. This unusual property arises from the fact that, as long as an algorithm (asymptotically) samples only keypoints, the increase in the $\zeta$ score will go to zero due to the fact that the confidence intervals shrink towards zero. So an algorithm could conceivably be zero-regret (in a policy regret sense) if it simply found (and endlessly sampled) one keypoint while ignoring all others. In contrast, this algorithm would not be zero-regret under an external regret definition (as we show formally in Theorem 3 and Corollary 2) since A would choose one of the low-sampled keypoints, and thus it would always be able to increase $\zeta$ by much more than the single-keypoint algorithm. Therefore, we adopt the external regret definition of Equation (52) in the rest of this paper.

We define a zero-regret algorithm as follows:

*An algorithm $B$ is **zero-regret** if and only if* $\lim_{T\to\infty} E\left[r_B\left(T\right)\right] = 0$

This problem setup is unusual: The typical bandit-style assumption is that the algorithm observes the reward/loss it receives after making its decision, and simply needs to keep that reward high enough relative to the optimal reward. In our case, however, the algorithm does not know the true keypoints, and yet regret is evaluated with respect to the true set of keypoints $K$. Therefore the true reward for each action is unobserved, making the problem more difficult.

## 5.1 Properties of zero-regret algorithms

To further illuminate this problem, we lay out some necessary properties of zero-regret algorithms. First, in Theorem 2 we show that an algorithm is zero-regret only if the proportion of samples assigned to non-keypoints converges to zero.

**Theorem 2.** *If* $|K| > 0$, $\lim_{T\to\infty} E\left[r_B\left(T\right)\right] = 0$ *only if:*

$$\lim_{T\to\infty} E\left[\frac{\sum_{i\notin K}|S_{i,T,B}|}{\sum_{i\in\mathcal{X}}|S_{i,T,B}|}\right] = 0$$

*Proof.* For a contradiction, assume there exists some algorithm $D$ such that

$$\lim_{T\to\infty} E\left[r_D\left(T\right) = 0\right] \text{ , but also } \lim_{T\to\infty} E\left[\frac{\sum_{i\notin K}|S_{T,i,D}|}{\sum_{i\in\mathcal{X}}|S_{T,i,D}|}\right] \neq 0 \tag{53}$$

Consider the regret on each $t'$ where $j_{t'} \notin K$. From equation (51) we have:

$$r_D\left(t'\right) = \left[\zeta\left(K, \mathbf{S_{t'-1,D}} \cup A\left(\mathbf{S_{t'-1,D}}\right)\right) - \zeta\left(K, \mathbf{S_{t'-1,D}}\right)\right] - \left[\zeta\left(K, \mathbf{S_{t',D}}\right) - \zeta\left(K, \mathbf{S_{t'-1,D}}\right)\right] \tag{54}$$

By Lemma 1 $\zeta$ will increase (or stay the same) after adding the sample chosen by $A$, since $A$ always selects keypoints. Therefore:

$$\left[\zeta\left(K, \mathbf{S_{t'-1,D}} \cup A\left(\mathbf{S_{t'-1,D}}\right)\right) - \zeta\left(K, \mathbf{S_{t'-1,D}}\right)\right] \geq 0 \tag{55}$$

Now, we know that on $t'$, algorithm $D$ selected a $j_t \notin K$. Therefore, we have:

$$\begin{aligned}
&\zeta\left(K, \mathbf{S_{t',D}}\right) - \zeta\left(K, \mathbf{S_{t'-1,D}}\right) \\
&= \frac{|K|}{\sum_{i\in K} C\left(|S_{i,t',D}|\right)} - \sum_{i\notin K}|S_{i,t',D}| - \frac{|K|}{\sum_{i\in K} C\left(|S_{i,t'-1,D}|\right)} - \sum_{i\notin K}|S_{i,t'-1,D}| && (56)\\
&= -\sum_{i\notin K}|S_{i,t',D}| + \sum_{i\notin K}|S_{i,t'-1,D}| && (57)\\
&= |S_{j_{t'},t'-1,D}| - |S_{j_{t'},t',D}| && (58)\\
&= -1 && (59)
\end{aligned}$$

Combining equations (55) and (59) into (54) gives us:

$$r_D\left(t'\right) \geq 0 - (-1) = 1 \tag{60}$$

So the regret on each $t'$ where $j_{t'} \notin K$ must be at least 1.

Now, if $E\left[\frac{\sum_{i \notin K} |S_{T,i,D}|}{\sum_{i \in \mathcal{X}} |S_{T,i,D}|}\right] = q_T$, then $E\left[r_D\left(T\right)\right] \geq Tq_T$, since each non-keypoint sample incurs at least 1 regret. And since $\lim_{T\to\infty} q_T \neq 0$, $E\left[r_D\left(T\right)\right]$ likewise cannot go to zero.

□

**Corollary 1.** *If $|K| > 0$, an algorithm which samples keypoints with probability $(1 - \epsilon_g)$ and non-keypoints (if there exist any) with probability $\epsilon_g$ cannot be zero regret.*

*Proof.* This follows immediately from Theorem 2, as if we assign probability $\epsilon_g$ to non-keypoints, $\lim_{T\to\infty} E\left[\frac{\sum_{i \notin K} |S_{i,T,B}|}{\sum_{i \in \mathcal{X}} |S_{i,T,B}|}\right] = \epsilon_g$. □

**Theorem 3.** *An algorithm $B$ is zero-regret only if, for any number $\ell$, for all $i \in K$, $lim_{T\to\infty} P(|S_{i,T,B}| \geq \ell) = 1$ .*

*Proof.* If $|K| = 0$, the condition trivially holds, so we consider $|K| > 0$. For a contradiction, assume there exists an algorithm $D$ such that $\lim_{T\to\infty} E\left[r_D\left(T\right)\right] = 0$ but for all $i \in K_n \subseteq K$, $P(|S_{i,T,D}| \geq \ell') < 1 - \delta'$ as $T \to \infty$ for some $\ell'$, where $|K_n| > 0$ and $\delta'$ is a fixed probability greater than 0. Consider cases where

$$|S_{i,T,D}| < \ell' \text{ for all } i \in K_n \text{ as } T \to \infty, \tag{61}$$

which must happen with probability at least $\delta'$. Note that if $D$ incurs at least expected regret $r$ for some $r > 0$ in these cases, it's overall expected regret will be at least $r\delta'$, so $D$ must be zero-regret in these cases as well.

First, if $K_n = K$, $\lim_{T\to\infty} E\left[\frac{\sum_{i \notin K} |S_{T,i,D}|}{\sum_{i \in \mathcal{X}} |S_{T,i,D}|}\right] = 1$ (since $A$ cannot continue sampling keypoints infinitely many times, it must shift an ever-increasing proportion of samples to non-keypoints). In this case, Theorem 2 tells us that $D$ cannot be zero-regret. So it must be the case that $K_n$ is a strict subset of $K$, and there exists some non-empty set $K_p = K - K_n$ of keypoints where for all $i \in K_p$, $|S_{i,T,D}| \to \infty$ as $T \to \infty$. In order to fulfill the condition in Theorem 2 (and given that $K_n$ cannot be sampled infinitely many times) it must also be the case that $\lim_{T\to\infty} E\left[\frac{\sum_{i \in \mathcal{X}-K_p} |S_{T,i,D}|}{\sum_{i \in \mathcal{X}} |S_{T,i,D}|}\right] = 0$, or conversely:

$$\lim_{T\to\infty} E\left[\frac{\sum_{i \in K_p} |S_{T,i,D}|}{\sum_{i \in \mathcal{X}} |S_{T,i,D}|}\right] = 1 \tag{62}$$

Since $D$ samples all $i \in K_p$ infinitely many times as $T \to \infty$, let $t'$ be a random variable denoting the first timestep such that

$$|S_{i,t',D}| > \ell' \text{ for all } i \in K_p \tag{63}$$

Now, consider the set of timesteps $\tau$ which contains all $T'$ such that the chosen x-value $j_{T',D} \in K_p$ and $T' > t'$. By equation (61) it must be the case that:

$$|S_{i,T',D}| < \ell' \text{ for all } i \in K_n \text{ and all } T' \in \tau \tag{64}$$

Now combining equations (63) and (64) gives us:

$$|S_{i_p,T',D}| > |S_{i_n,T',D}| \text{ for all } i_p \in K_p, i_n \in K_n \text{ and all } T' \in \tau \tag{65}$$

Now, since $|S_{i_p,T',D}| \to \infty$ for all $i_p \in K_p$, it must be true that, for all $i_p \in K_p$:

$$\lim_{T'\to\infty} \sqrt{\frac{ln\left(\frac{2}{\delta}\right)}{2|S_{i_p,T',D}| + \epsilon}} = 0 \tag{66}$$

$$\lim_{T'\to\infty} C(|S_{i_p,T',D}|) = 0 \tag{67}$$

$$\lim_{T'\to\infty} C(|S_{i_p,T',D}|) - C(|S_{i_p,T'-1,D}|) = 0 \text{ for all } i_p \in K_p \tag{68}$$

Now, observe that, since $j_{T',D} \in K_p$,

$$C(|S_{i_n,T',D}|) = C(|S_{i_n,T'-1,D}|) \text{ for all } i_n \in K_n \tag{69}$$

Combining equations (68) and (69) gives us

$$\lim_{T'\to\infty} \sum_{i_p\in K_p} C(|S_{i_p,T',D}|) - C(|S_{i_p,T'-1,D}|) + \sum_{i_n\in K_n} C(|S_{i_n,T',D}|) - C(|S_{i_n,T'-1,D}|) = 0 \tag{70}$$

$$\lim_{T'\to\infty} \sum_{i\in K} C(|S_{i,T',D}|) - \sum_{i\in K} C(|S_{i,T'-1,D}|) = 0 \tag{71}$$

$$\lim_{T'\to\infty} d(\mathbf{S_{T',D}}) - d(\mathbf{S_{T'-1,D}}) = 0 \tag{72}$$

Note that while the difference in equation (72) goes to zero, $d(\mathbf{S_{T',D}})$ itself does not go to zero, since for each $i \in K_n$, $C(|S_{i_n,T',D}|) \geq \sum_{i_n\in K_n} C(\ell')$ by Lemma 2 and equation (61). Therefore:

$$\lim_{T'\to\infty} \frac{|K|}{d(\mathbf{S_{T',D}})} - \frac{|K|}{d(\mathbf{S_{T'-1,D}})} = 0 \tag{73}$$

$$\lim_{T'\to\infty} \zeta(K, \mathbf{S_{T',D}}) - \zeta(K, \mathbf{S_{T'-1,D}}) = 0 \tag{74}$$

Where the last line follows since $j_{T',D} \in K$ for all $T' \in \tau$.

Let $i_{A,T'} = i_{A(S_{T'-1,D})}$ be the x-value $A$ would choose given the set $S_{T'-1,D}$. $A$ selects $\operatorname{argmin}_{i\in K} |S_i|$ by definition, so by equation (65), $i_A \in K_n$ and thus by equation (61):

$$|S_{i_A,T',D}| \leq \ell' \tag{75}$$

Now examine the difference in $\zeta$ before and after adding the sample from $A$. Since $A$ always selects keypoints we have:

$$\begin{aligned}&\zeta\left(K, \mathbf{S_{T'-1}} \cup A\left(\mathbf{S_{T'-1}}\right)\right) - \zeta\left(K, \mathbf{S_{T'-1}}\right) \\ &\quad= \frac{|K|}{C(|S_{i_A,T'-1}|+1) + \sum_{i\in K, i\neq i_A} C(|S_{i,T'-1}|)} - \frac{|K|}{C(|S_{i_A,T'-1}|) + \sum_{i\in K, i\neq i_A} C(|S_{i_A,T'-1}|)}\end{aligned} \tag{76}$$

We wish to lower bound this difference. Since $\frac{|K|}{x}$ is convex and decreasing if $x > 0$, to minimize the difference we make $\frac{|K|}{x}$ as large as possible, which means making the denominator $x$ as small as possible. Since the function $C$ is always non-negative:

$$\zeta\left(K, \mathbf{S_{T'-1}} \cup A\left(\mathbf{S_{T'-1}}\right)\right) - \zeta\left(K, \mathbf{S_{T'-1}}\right) \geq \frac{|K|}{C(|S_{i_A,T'-1}|+1)} - \frac{|K|}{C(|S_{i_A,T'-1}|)} \tag{77}$$

$$\geq \frac{|\mathcal{X}|}{C(|S_{i_A,T'-1}|+1)} - \frac{|\mathcal{X}|}{C(|S_{i_A,T'-1}|)} \tag{78}$$

$$\text{Let } \zeta_{min} = \min_{x\in[0,\ell']} \frac{|\mathcal{X}|}{C(x+1)} - \frac{|\mathcal{X}|}{C(x)} \tag{79}$$

$\zeta_{min}$ is a constant, which is always greater than zero since $C$ monotonically decreases (see Lemma 2). Now, combining equations (79) and (78):

$$\zeta\left(K, \mathbf{S_{T'-1}} \cup A\left(\mathbf{S_{T'-1}}\right)\right) - \zeta\left(K, \mathbf{S_{T'-1}}\right) \geq \zeta_{min} \tag{80}$$

By the definition of regret in equation (51) we have:

$$r_D\left(T'\right) \geq \zeta_{min} - \zeta(K, \mathbf{S_{T',D}}) - \zeta(K, \mathbf{S_{T'-1,D}}) \tag{81}$$

However, recall that this holds only for $T' \in \tau$. In general, we can write:

$$E\left[r_D\left(T\right)\right] \geq P(T \in \tau)\left(\zeta_{min} - E\left[\zeta(K, \mathbf{S_{T,D}}) - \zeta(K, \mathbf{S_{T-1,D}})\right]\right) \tag{82}$$

From equation (74), the last two terms go to zero in the limit, giving us:

$$\lim_{T\to\infty} E\left[r_D\left(T\right)\right] \geq P(T \in \tau)\left(\zeta_{min} - 0\right) \tag{83}$$

Now, equation (62) (along with the definition of $\tau$) implies that $\lim_{T\to\infty} P(T \in \tau) = 1$. However, note that equation (62) is derived from equation (61), which holds only with probability $\delta'$. So

$$\lim_{T\to\infty} E\left[r_D\left(T\right)\right] \geq \delta'\zeta_{min} \tag{84}$$

And since $\zeta_{min}$ and $\delta'$ are both constants greater than 0, $D$ cannot be zero regret.

□

**Corollary 2.** *An algorithm which samples round-robin until the first keypoint is found, then samples that keypoint forever (perhaps reverting to round-robin if that keypoint is deleted), is not zero regret if $|K| > 1$.*

*Proof.* This follows directly from Theorem 3: if we have $\mathcal{H}$ such that the first true keypoint $i_c$ is revealed after $\ell_c$ samples, and not removed thereafter, and for all $j \in K$ such that $j \neq i_c$, more than $\ell_c + 1$ samples are required to identify it as a keypoint, then
$P(|S_{j,t}| > \ell_c + 1) = 0$ for $j \in K, j \neq i_c$ and all $t$. □

# 6 A Zero-Regret Bayesian Algorithm

In Algorithm 1 we present our Threshold Estimating Sampling Algorithm (TESA). Although some parts of TESA are similar to epsilon-greedy, the Bayesian threshold-learning procedure (explained further in the main text) improves performance. Specifically, although epsilon-greedy is not zero-regret (Corollary 1), TESA is a zero-regret algorithm, as we show in Theorem 4.

**Algorithm 1** Threshold Estimating Sampling Algorithm (TESA)

Input: Tradeoff parameter $\epsilon_p \in [0, 1)$, prior scale $b_0 > 0$, prior shape $a > 0$
$\hat{K} \leftarrow \emptyset; \forall_{i\in\mathcal{X}} S_{i,0} \leftarrow \emptyset, S_{i,1} \leftarrow \emptyset$
**for** t = 1 to $T$ **do**
  **if** $t \mod \mathcal{I} = 0$ **then**
    Send $(f_h, f_l, \hat{f})$ to human $\mathcal{H}$, get keypoints $\hat{K}_t$
    **for all** $i \in \hat{K}_t - \hat{K}$ **do**
      $b_i \leftarrow \frac{|S_{i,t}|+|S_{i,t-\mathcal{I}}|}{2}$
    $\hat{K} \leftarrow \hat{K}_t$
  $x_m \leftarrow \max\left(b_0, \max_{i\in\hat{K}} b_i\right)$
  $h \sim \text{Pareto}\left(x_m, a + |\hat{K}|\right)$
  $i_k \leftarrow \text{argmin}_{i\in\hat{K}} |S_{i,t}|$
  $i_h \leftarrow \text{argmin}_{i\in\mathcal{X}} |S_{i,t}|$
  $r \sim \text{Uniform}\,(0, 1)$
  **if** $|\hat{K}| > 0$ **and** $(r \geq \epsilon_p$ **or** $|S_{i_h,t}| \geq h)$ **then**
    $i_c \leftarrow i_k$
  **else**
    $i_c \leftarrow i_h$
  Choose x value $i_c$; Receive new sample $y_t$
  $S_{i_c,t+1} \leftarrow S_{i_c,t} \cup \{y_t\}$
  $\forall_{i\in\mathcal{X} \text{ s. t. } i\neq i_c} S_{i,t+1} \leftarrow S_{i,t}$

**Theorem 4.** *TESA (Algorithm 1) is zero regret.*

*Proof.* For convenience in this proof, when we leave off the algorithm-specific subscripts, it refers to TESA. For instance, $S_{i,t}$ refers to the set of samples TESA would have for x-value $i$ at timestep $t$.

First, if $K = \emptyset$ then $\zeta$ is the same for every algorithm and the theorem is trivial, so we here consider cases where $|K| > 0$.

Let $X_p \subseteq \mathcal{X}$ be the set of all $i \in \mathcal{X}$ such that $|S_{i,t}| \to \infty$ as $t \to \infty$. Let $X_n = \mathcal{X} - X_p$. For each $i \in X_n$, it must be the case that $P(|S_{i,T,C}| \geq \ell') < 1 - \delta'$ as $T \to \infty$ for some $\ell'$ and fixed $\delta' > 0$. Therefore, consider cases where $P(|S_{i,T,C}| < \ell')$ for all $i \in X_n$ as $T \to \infty$, which happens with fixed probability $\delta'$.

Now, all $i \in X_p$ will be sampled infinitely many times, so consider $t''$ sufficiently large such that $|S_{i,t'',C}| \leq \ell'$ for all $i \in X_n$ but, for all $i \in X_p$, $|S_{i,t'',C}| > \ell'$. Then, at timesteps $t \geq t''$, consider two (non-exhaustive) cases, $|X_n \cap \hat{K}_t| > 0$ and $|X_n \cap \hat{K}_t| = 0, |X_n| > 0$.

**Case 1:** $|X_n \cap \hat{K}_t| > 0$.

In this case consider any $j'$ such that $j' \in X_n$ and $j' \in \hat{K}_t$. Since $\hat{K}_t$ is non-empty, with at least probability $(1 - \epsilon_p)$ Algorithm 1 will sample a keypoint. Since Algorithm 1 always samples the keypoint with fewest samples, it will chose $j'$ (with has at most $\ell'$ samples) over any $i \in X_p$ (which has more than $\ell'$ samples at timestep $t \geq t''$). Since there is a constant probability $(1 - \epsilon_p)$) of sampling any $x \in X_n$ at timesteps $t \geq t''$, they would eventually have more than $\ell'$ samples almost surely, therefore this case has probability zero.

**Case 2:** $|X_n \cap \hat{K}_t| = 0, |X_n| > 0$

The probability that Algorithm 1 will sample non-keypoint is as follows:

$$P(j_t \notin \hat{K}_t) = (P(|\hat{K}_t| = 0) + P(r < \epsilon_p)P(|S_{i_h,t}| < h)) \tag{85}$$

$$\geq \epsilon_p P(|S_{i_h,t}| < h) \tag{86}$$

Since Algorithm 1 sets $i_h$ to be the non-keypoint with minimum samples, when it chooses a non-keypoint at timstep $t \geq t''$ it will always choose $i_h \in X_n$ (as each element of $X_n$ has at most $\ell'$ samples) over any $i \in X_p$ (which has more than $\ell'$ samples at timestep $t \geq t''$). So

$$P(j_t \in X_n) \geq \epsilon_p P(|S_{i_h,t}| < h) \tag{87}$$

$$\geq \epsilon_p P(\ell' < h) = \epsilon_p P(h > \ell') \tag{88}$$

Note that $h$ is distributed as Pareto $\left(\max\left(b_0, \max_{i \in \hat{K}_t} b_i\right), a + |\hat{K}_t|\right)$, which is drawn independently on each timestep. So by the Pareto complementary CDF and equation (88):

$$P(j_t \in X_n) \geq \epsilon_p \left(\frac{x_m}{\ell'}\right)^{|\hat{K}_t|+a} \tag{89}$$

$$\geq \epsilon_p \left(\frac{\max(b_0, \max_{i \in \hat{K}_t} b_i)}{\ell'}\right)^{|\hat{K}_t|+a} \tag{90}$$

$$\geq \min_{a' \in [0, |\mathcal{X}|]} \epsilon_p \left(\frac{b_0}{\ell'}\right)^{a'+a} \tag{91}$$

The total probability of sampling some $i \in X_n$ is lower bounded by a constant per equation (91) at timesteps $t \geq t''$, therefore the number of samples for at least one $i \in X_n$ would eventually become greater than $\ell'$ almost surely. So this case must have probability zero.

Note that Case 1 and Case 2 each have probability zero. The only case not covered by case 1 and 2 is $|X_n| = 0$, so $P(|X_n| = 0) = 1$, meaning

$$\text{For all x-values } i \in \mathcal{X}, |S_{i,T,C}| \to \infty \text{ as } T \to \infty, \tag{92}$$

Therefore, for any $\delta''$ there exists a sufficiently large $t'$ such that $|S_{i,t'-1}| > n_i$ for all $i \in \mathcal{X}$ with probability $1 - \delta''$. Now if $|S_{i,t'-1}| > n_i$ for all $i \in \mathcal{X}$, $\hat{K}_t = K$ for all $t \geq t' - 1$ by Assumption 1. Therefore:

$$\text{For sufficiently large } t, \hat{K}_t = K \text{ under Algorithm1 almost surely} \tag{93}$$

Now the expected $\zeta$-score in this case will be:

$$E\left[\zeta\left(K, \mathbf{S_t}\right)\right] = E\left[\frac{|K|}{\sum_{i \in K} C_i\left(|S_{i,t}|\right)} - \sum_{i \notin K} |S_{i,t}|\right] \tag{94}$$

$$= E\left[\frac{|\hat{K}_t|}{\sum_{i \in \hat{K}_t} C_i\left(|S_{i,t}|\right)} - \sum_{i \notin \hat{K}_t} |S_{i,t}|\right] \tag{95}$$

And since $\hat{K}_{t-1} = K$ as well, the difference in $\zeta$ score between timesteps $t-1$ and $t$ is:

$$
\begin{aligned}
&E\left[\zeta\left(K, \mathbf{S_t}\right) - \zeta\left(K, \mathbf{S_{t-1}}\right)\right] \\
=&E\left[\frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \sum_{i \notin \hat{K}_t} |S_{i,t}| - \left(\frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})} - \sum_{i \notin \hat{K}_t} |S_{i,t-1}|\right)\right] && (96)\\
=&E\left[\frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})}\right] - E\left[\sum_{i \notin \hat{K}_t} |S_{i,t}| - |S_{i,t-1}|\right] && (97)\\
=&E\left[\left(\frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})}\right)\right] - P(j_t \notin \hat{K}_t) E\left[(|S_{j_t,t-1}| + 1) - |S_{j_t,t-1}|\right] && (98)\\
=&E\left[\left(\frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})}\right)\right] - P(j_t \notin \hat{K}_t) && (99)\\
=&\left(1 - P(j_t \notin \hat{K}_t)\right) E\left[\left(\frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})}\right) \middle| j_t \in \hat{K}_t\right] - P(j_t \notin \hat{K}_t) && (100)
\end{aligned}
$$

Where the last line follows since $\left(\frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})}\right) = 0$ if $P(j_t \notin \hat{K}_t)$.

Now, examine the rightmost term in equation (100):

$$
\begin{aligned}
P(j_t \notin \hat{K}_t) =& P(|\hat{K}_t| = 0) + P(r < \epsilon_p) P(|S_{i_h,t}| < h) && (101)\\
=& P(|K| = 0) + \epsilon_p P(|S_{i_h,t}| < h) && (102)\\
=& \epsilon_p P(|S_{i_h,t}| < h) && (103)
\end{aligned}
$$

Where the last equation follows since we assumed $|K| > 0$. Combining equations (103) and (100) gives us:

$$
\begin{aligned}
&E\left[\zeta\left(K, \mathbf{S_t}\right) - \zeta\left(K, \mathbf{S_{t-1}}\right)\right] \\
&\quad \geq (1 - \epsilon_p P(|S_{i_h,t}| < h)) E\left[\left(\frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})}\right) \middle| j_t \in \hat{K}_t\right] - \epsilon_p P(|S_{i_h,t}| < h) && (104)
\end{aligned}
$$

Recall that for any $\delta''$ there exists a sufficiently large $t'$ that with probability $(1 - \delta'')$ equation (104) holds for all $t \geq t'$, so:

$$
\begin{aligned}
&\lim_{t \to \infty} E\left[\zeta\left(K, \mathbf{S_t}\right) - \zeta\left(K, \mathbf{S_{t-1}}\right)\right] \\
&\geq \lim_{t \to \infty} \left((1 - \epsilon_p P(|S_{i_h,t}| < h)) E\left[\left(\frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})}\right) \middle| j_t \in \hat{K}_t\right]\right) - \lim_{t \to \infty} \epsilon_p P(|S_{i_h,t}| < h) && (105)
\end{aligned}
$$

Now, examine the last term in equation (105). Since $h$ is drawn from a Pareto distribution on each timestep using the Pareto complementary CDF:

$$
\begin{aligned}
\lim_{t \to \infty} \epsilon_p P(|S_{i_h,t}| < h) &= \lim_{t \to \infty} \epsilon_p \left(\frac{\max(b_0, \max_{i \in \hat{K}_t} b_i)}{|S_{i_h,t-1}|}\right)^{|\hat{K}_t| + a} && (106)\\
&\leq \lim_{t \to \infty} \epsilon_p \left(\frac{\max(b_0, \max_{i \in K} n_i + \frac{\mathcal{I}}{2})}{|S_{i_h,t-1}|}\right)^{|\hat{K}_t| + a} && (107)
\end{aligned}
$$

Where the last line follows since the maximum number of samples it can take to identify that something is (or is not) a keypoint is $\max_{i \in \mathcal{X}} n_i$ by Assumption 1, and then we devote at most $\mathcal{I}$ more samples to it (with the midpoint equation on line 7 causing it to be at most $\frac{\mathcal{I}}{2}$ off). Now, using the fact in equation (93):

$$\lim_{t\to\infty} \epsilon_p P(|S_{i_h,t}| < h) \leq \lim_{t\to\infty} \epsilon_p \left( \frac{\max(b_0, \max_{i\in K} n_i + \frac{\mathcal{I}}{2})}{|S_{i_h,t-1}|} \right)^{|K|+a} \tag{108}$$

Now, recall that we established in equation (92) that as $t \to \infty$, $|S_{i,t-1}| \to \infty$ for all $i$. Since the other terms in equation (108) are constants:

$$\lim_{t\to\infty} \epsilon_p P(|S_{i_h,t}| < h) = 0 \tag{109}$$

Now, combining equations (105) and (109) gives us that:

$$\lim_{t\to\infty} E\left[\zeta\left(K, \mathbf{S_t}\right) - \zeta\left(K, \mathbf{S_{t-1}}\right)\right] \geq \lim_{t\to\infty} \left( E\left[ \left( \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})} \right) \middle| j_t \in \hat{K}_t \right] \right) \tag{110}$$

Now, combining equation (110) with our definition of regret (equation (51)) gives us:

$$\lim_{t\to\infty} E\left[r\left(t\right)\right] \leq \lim_{t\to\infty} \left( E\left[\zeta\left(K, \mathbf{S_{t-1}} \cup A\left(\mathbf{S_{t-1}}\right)\right) - \zeta\left(K, \mathbf{S_{t-1}}\right)\right] - E\left[ \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})} \middle| j_t \in \hat{K}_t \right] \right) \tag{111}$$

Now examining the left two terms of the above equation gives us:

$$\begin{aligned}
&E\left[\zeta\left(K, \mathbf{S_{t-1}} \cup A\left(\mathbf{S_{t-1}}\right)\right) - \zeta\left(K, \mathbf{S_{t-1}}\right)\right] \\
&= E\left[ \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}} \cup A\left(\mathbf{S_{t-1}}\right))} - \sum_{i\notin\hat{K}_t} |S_{i,t}| - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})} - \sum_{i\notin\hat{K}_t} |S_{i,t-1}| \right] &(112)\\
&= E\left[ \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}} \cup A\left(\mathbf{S_{t-1}}\right))} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})} \right] &(113)
\end{aligned}$$

Where the last line follows since $A$ never selects non-keypoints. Combining equations (113) and (111) gives us:

$$\begin{aligned}
&\lim_{t\to\infty} E\left[r\left(t\right)\right] \\
&\leq \lim_{t\to\infty} \left( E\left[ \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}} \cup A\left(\mathbf{S_{t-1}}\right))} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})} \right] - E\left[ \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_t})} - \frac{|\hat{K}_t|}{d(\hat{K}_t, \mathbf{S_{t-1}})} \middle| j_t \in \hat{K}_t \right] \right) &(114)
\end{aligned}$$

Now, $A\left(\mathbf{S_{t-1}}\right) = \text{argmin}_{i\in K} |S_{i,t-1}|$ by definition. By equation (93), for sufficiently large $t$, $A\left(\mathbf{S_{t-1}}\right) = \text{argmin}_{i\in\hat{K}_t} |S_{i,t-1}|$ almost surely. And if $j_t \in \hat{K}_t$, Algorithm 1 selects $j_t = i_k = \text{argmin}_{i\in\hat{K}_t} |S_{i,t-1}|$. Therefore, for sufficiently large $t$ the terms in equation (114) cancel, leaving us with:

$$\lim_{t\to\infty} E\left[r\left(t\right)\right] \leq 0 \tag{115}$$

Therefore, Algorithm 1 is zero regret. □

# 7 User Study Details

Here we provide additional details on the user study - the descriptions of all three environments and the associated quizzes, the keypoint quizzes and instructions, and screenshots of the more complex parts of the task.

## 7.1 Economics Domain

*Economics Description*

Avocados are small, green, hand-sized fruits that are growing in popularity around the world and are normally sold at many supermarkets across the United States.

The retail price of avocados is key to the number of avocados sold. Economic theory states that the higher the price, the less likely people are to buy something. As an example, imagine you are presented with 2 identical avocados to purchase, one priced at 1 dollar and the other at 100 dollars. Which would you choose?

With all that being said, economic theory is generally only good in a big picture sense, so in this HIT you will run an experiment varying the price of Hass Avocados and observing how the number sold changes. You will see a graph containing the price on the horizontal (x) axis and the number sold on the vertical (y) axis.

*Economics Quiz*

1. Other things being equal, if the price goes up, what happens to sales?

- A. Sales go down.
- B. Sales stay the same.
- C. Sales go up.
- D. None of the above

2. At which location is a person most likely to find avocados?

- Antarctic Ocean
- Grocery shop
- Library
- Casino

## 7.2 Cognitive Psychology Domain

*Cognitive Psychology Description*

Scientists are interested in understanding peoples reaction times via experimentation. Reaction time is a stop-watch measurement of things like how long it takes a sprinter to start moving after a referee fires a starting pistol. An experiment is going to be conducted to measure peoples reaction times. The participants in this experiment will be college students, with varying majors, in Germany. They will have no prior experience with this experiment. In this experiment, participants will sit in front of a computer screen in a laboratory. They will be shown a series of square shapes, one square at a time. For each square, the participant will guess the size of the square as quickly as possible.

Scientists hypothesize that how similar in size the current square is to the previous square will affect the participants reaction time.

Imagine, if you will, viewing a square, and guessing its size. Then, you would view another square of the same size and guess its size. Maybe it wouldnt take long to guess the familiar size. But if the two square sizes were very different (for example, a really big and really small square), it might take much longer.

We could visualize the reaction time as a graph. The reaction time would be on the vertical (y) axis, and the horizontal (x) axis would have the square size difference, ranging from -18mm (meaning the current square is 18mm smaller than the previous square) to 18 mm (meaning the current square is 18mm bigger than the previous square). In this HIT, you will play the role of a scientist conducting an experiment to determine the shape of this graph. In general, psychologists would expect this graph to look almost exactly like a capital letter V (Petzold, P., Haubensak, G., 2004), with the base of the ”V” at 0mm, but it may or may not be the case here.

*Cognitive Psychology Quiz*

1. This experiment is designed to test whether reaction time is connected to:

- A. The color of two squares
- B. The participants reaction time

- C. The size difference between consecutive squares
- D. Internet connection speed

2. Psychologists generally believe the reaction time will be smallest when:

- A. One square is much bigger than the next.
- B. Two consecutive squares are about the same size.
- C. The squares are of two very different sizes
- D. The squares are of the smallest possible size.

## 7.3 Mental Health Domain

*Mental Health Description*

Mental illness has become an increasingly large topic of discussion in the modern age, especially within the tech industry.

Mood disorders are mental illnesses that involve persistent distorted emotional functioning. People with mood disorders experience emotional states that are inconsistent with their circumstances, for example abnormally strong and persistent feelings of anxiety, fear, sadness, or depression. In bipolar disorder, people sometimes experience periods of abnormally elevated mood. The most common mood disorders involve persistent elevated negative emotions such as depression and anxiety.

The percentage of the population with a mood disorder tends to decrease with age, with the highest rates of mood disorders occurring among young adults (National Institute of Mental Health, 2017).

In this HIT, you will run an experiment looking deeper at the question of how age (on the horizontal x-axis) affects the percentage of mood disorder within the tech industry (on the vertical y-axis). Note that individuals within the tech industry may or may not behave the same way as the general population.

*Mental Health Quiz*

1. How do mood disorders typically relate to age, according to the National Institute of Mental Health?

- A. Middle aged people suffer from mood disorders at a much higher rate than any other age group.
- B. Younger adults suffer from mood disorders at a much higher rate than any other age group.
- C. Older people suffer from mood disorders at a much higher rate than any other age group.
- D. Younger adults suffer from mood disorders at a much lower rate than any other age group.

2. Which of the following best defines a mood disorder?

- A. A type of mental illness that mostly affects your discussion ability.
- B. Mismatching colors that ruin a vibe.
- C. A type of mental illness that causes people to experience very mild happiness about the tech industry.
- D. A type of mental illness that causes abnormal emotions.

## 7.4 Understanding Keypoints

*Keypoint Description*

Recall that you are playing the role of a scientist running an experiment. Collecting data is time intensive and expensive, so you want to be careful to collect the most useful data possible. To help the system collect better data, in a later part of the task you will identify scientifically interesting points, called **key points**, on the graph. These points will be used to guide data collection.

An scientifically interesting point (key point) is a place on the graph that is unexpected. This is a bit hard to define precisely as many possible things could be scientifically interesting, but we'll give an example.

For a hypothetical example, imagine a graph of temperature that generally shows a 0 degrees Celsius (32 degrees Fahrenheit) winter. But on December 1st, it was a balmy 23 degrees Celsius (73 degrees Fahrenheit). Then December 1st could be a scientifically interesting point, but it could also be simply the result of a malfunctioning thermometer. But if you had a lot of data for December 1st agreeing with the high temperature, it would be very interesting.

The data collected so far will be visualized as shown to the right. The dashed purple line indicates the best guess as to where the data lies, but it could actually be anywhere in the blue area (the blue area essentially represents confidence intervals). **The larger the vertical blue area, the *less* data we have at that x-value.**

**A key point will be visualized with a star. In a few screens, your task will be to add these keypoints and then drag and drop them onto the scientifically interesting points of the graph.** [Picture of star]

**Be careful! You should NOT mark any keypoints if we haven't collected enough datapoints yet!** A very large blue area means that we don't have many datapoints at that location, and the total number of datapoints (N) used to build the graph is usually shown near the bottom.

*Keypoint Quizzes*

1. Users are shown Figure 1 and asked: Examine the below graph. At which x-value is there the most data?

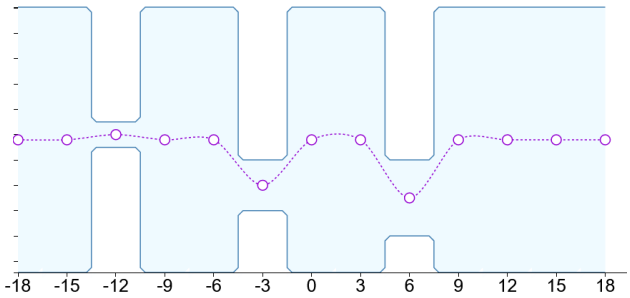


Figure 1: The figure for the first keypoint question.

- A. 15
- B. 6
- C. -3
- D. -12

2. Users are shown Figure 2 and asked: How many key points should you mark in the graph, below? Recall that key points are **scientifically interesting points at which we have enough data** to be somewhat confident in their values.

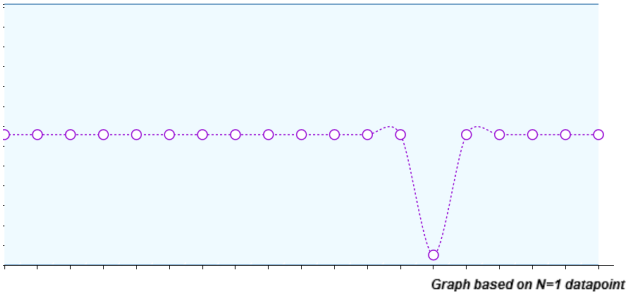


Figure 2: The figure for the second keypoint question.

- A. I should mark 0 key points.
- B. I should mark 1 key points.
- C. I should mark 2 key points.
- D. I should mark 5 key points.
- E. I should mark 13 key points.

## 7.5 Task Screenshots

Figures 3- 5 show screenshots of the more visually complicated parts of the task.

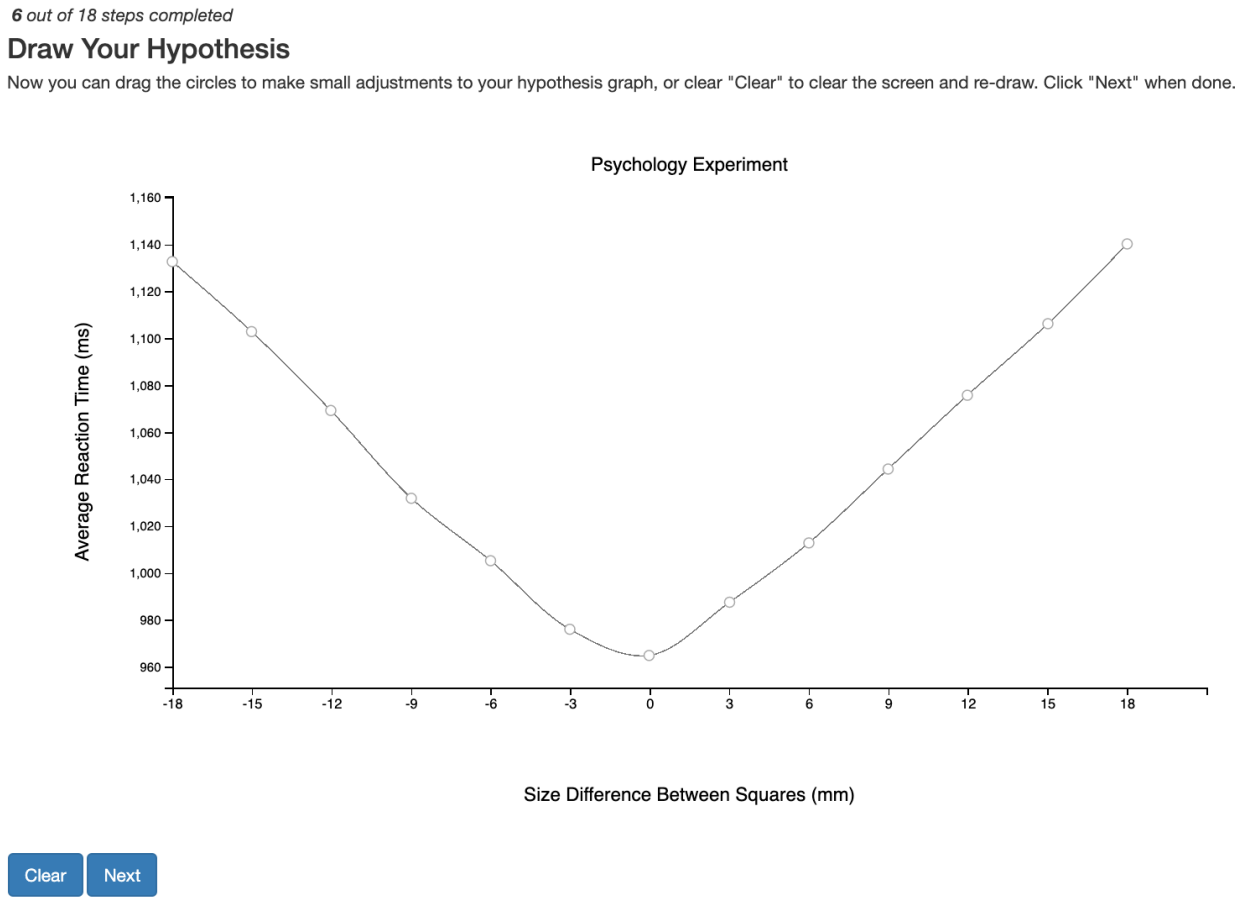


Figure 3: The user interface for the prior specification part of the experiment. Participants draw a curve in freehand, then make smaller adjustments the curve via vertically dragging points (represented by circles) on the curve.

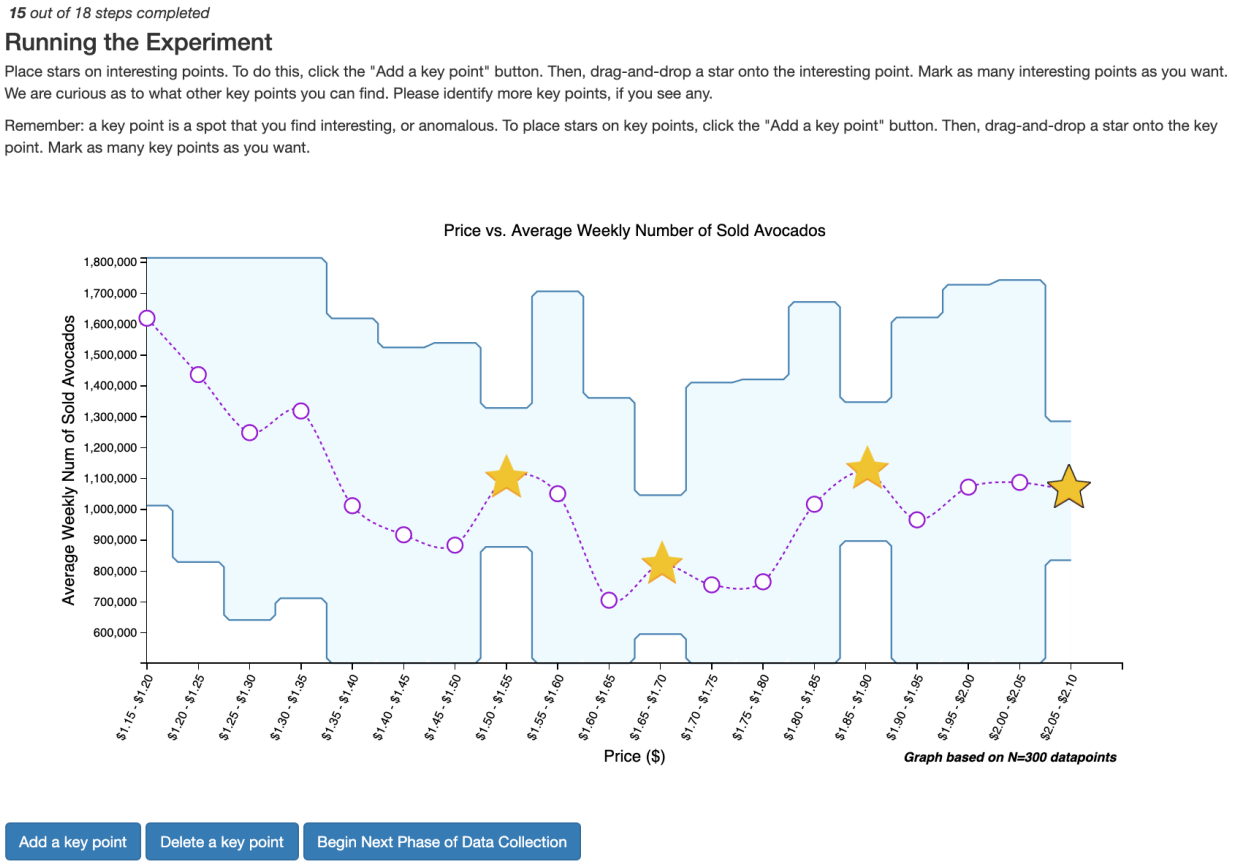


Figure 4: Once users have specified their prior curve, they move on to several iterations of our keypoint selection interface. We visualize the current estimated function (purple dotted line) and confidence region (blue area). Users can add and delete keypoints as well as drag them to interesting regions of the function.

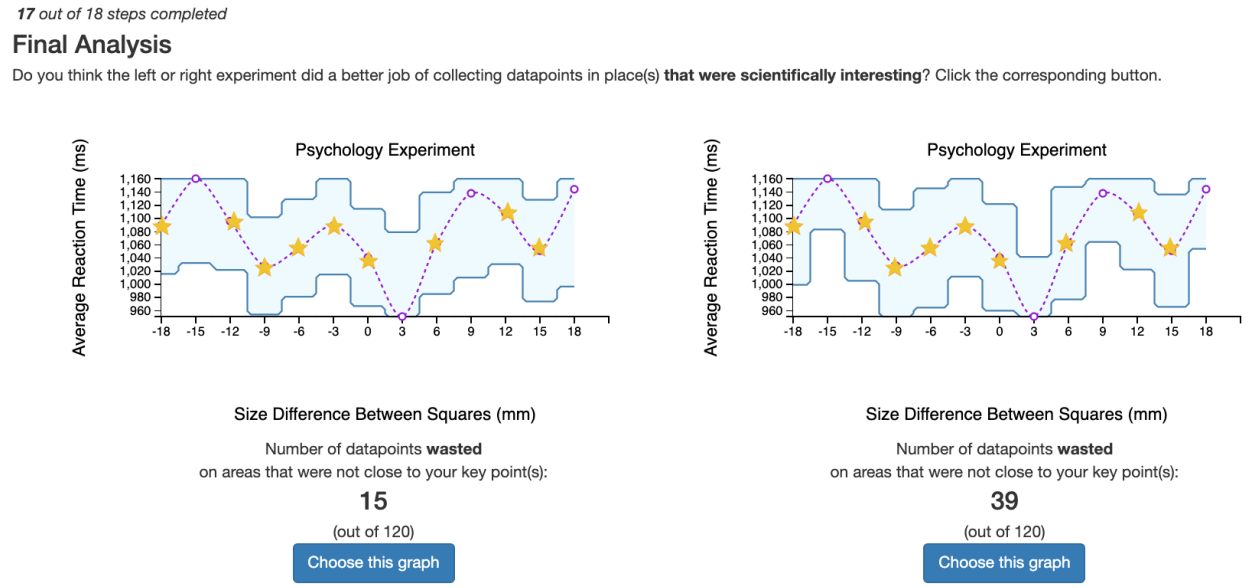


Figure 5: Once users have specified several iterations of keypoints, they move on to the final phase. Participants choose which graph they think did the best job collecting data points in places they deem scientifically interesting. One of these graphs is generated by UWPS based on the prior curve while the other is generated by TESA. Their position is randomized and users are not told which is which.

# References

[1] R. Arora, O. Dekel, and A. Tewari, "Online bandit learning against an adaptive adversary: from regret to policy regret," in *Proceedings of the 29th International Conference on Machine Learning*, pp. 1747–1754, 2012.

[2] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM Journal on Computing*, vol. 32, no. 1, pp. 48–77, 2002.