FUNDAMENTA
MATHEMATICAE
220 (2013)

Universality of the μ-predictor

by

Christopher S. Hardin (New York)

Abstract. For suitable topological spaces X and Y, given a continuous function $f : X \to Y$ and a point $x \in X$, one can determine the value of $f(x)$ from the values of f on a deleted neighborhood of x by taking the limit of f. If f is not required to be continuous, it is impossible to determine $f(x)$ from this information (provided $|Y| \geq 2$), but as the author and Alan Taylor showed in 2009, there is nevertheless a means of guessing $f(x)$, called the *μ-predictor*, that will be correct except on a small set; specifically, if X is T_0, then the guesses will be correct except on a scattered set. In this paper, we show that, when X is T_0, every predictor that performs this well is a special case of the μ-predictor.

1. Introduction

1.1. Background. Before discussing predictors, we must establish some terminology and notation.

Given a topological space X, a *deleted neighborhood* of a point $x \in X$ is a set $V - \{x\}$ where V is an open set containing x. A set $A \subseteq X$ is *scattered* if every nonempty $A' \subseteq A$ has isolated points; that is, there exists $x \in A'$ with neighborhood V such that $A' \cap V = \{x\}$ (in which case we say *x is isolated in A'* and *V isolates x in A'*). Weakening $A' \cap V = \{x\}$ to $A' \cap V$ being finite, we get the notion of *weakly scattered* (and weak isolation). Morgan refers to weakly scattered sets as *separated sets* [Mor90]. Scattered sets are always weakly scattered; in T_0 spaces, the two notions are equivalent. Given any partially ordered set P, the *downward* (resp. *upward*) *topology* on P has as open sets those that are downward (resp. upward) closed. The upward or downward topology is always T_0 but not T_1 unless all points are incomparable to each other. Given a topology $\mathcal{U}$ on X, we define the preorder $\leq$ on X by $x \leq y$ iff every neighborhood of y contains x. Note that $\leq$ is a partial order iff $\mathcal{U}$ is T_0 (and is trivial iff $\mathcal{U}$ is T_1). Also, if $\mathcal{U}$ is the downward topology on some partial order, then $\leq$ coincides with that

2010 *Mathematics Subject Classification*: Primary 03E05; Secondary 54H99.
Key words and phrases: sets, mu-predictor, fixed points.

DOI: 10.4064/fm220-3-4

order. We denote the closure of a set $A \subseteq X$ by $\overline{A}$, and write $\overline{x}$ as shorthand for $\overline{\{x\}}$. Note that $\overline{x} = \{y \in X \mid x \leq y\}$.

The set of functions from X to Y is denoted ${}^X Y$; our convention throughout the paper is that X is a topological space and Y is a set. Given $f \in {}^X Y$, $a \in X$, $b \in Y$, we use $f[b/a]$ to denote the function $f' \in {}^X Y$ with $f'(a) = b$ and $f'(x) = f(x)$ for $x \neq a$. For $f, g \in {}^X Y$, we abuse the notation for symmetric difference slightly and let $f \triangle g = \{x \in X \mid f(x) \neq g(x)\}$. For each $x \in X$, we define the equivalence relation $\approx_x$ on ${}^X Y$ by $f \approx_x g$ iff x has a deleted neighborhood Q such that $(f \triangle g) \cap Q = \emptyset$; informally, $f \approx_x g$ means that f and g are equal near (but not necessarily at) x. The $\approx_x$-equivalence class of f is denoted $[f]_x$, and is also called the *deleted germ of f at x*. If $S : {}^X Y \to {}^X Y$, we use Sf as shorthand for $S(f)$.

1.2. Predictors. We are concerned with the problem of, at each x, guessing the value of $f(x)$ from $[f]_x$. A *predictor (for ${}^X Y$)* is a function $S : {}^X Y \to {}^X Y$ such that for all $x \in X$,

$$f \approx_x g \ \Rightarrow \ Sf(x) = Sg(x). \tag{1.1}$$

We think of $Sf(x)$ as a guess for the value of $f(x)$; formula (1.1) requires that this guess is based only on the values of f near (but not at) x. Observe that limits, when they exist, behave this way: if $\lim_{t \to x} f(t)$ exists and $f \approx_x g$, then $\lim_{t \to x} g(t) = \lim_{t \to x} f(t)$. We say that *$S$ guesses f correctly at x* if $Sf(x) = f(x)$; accordingly, $Sf \triangle f$ is the set of points where S guesses f incorrectly. For $\mathcal{I} \subseteq \mathcal{P}(X)$ (typically an ideal), we call S an *$\mathcal{I}$-error* predictor if $Sf \triangle f \in \mathcal{I}$ for all $f \in {}^X Y$; in the special cases where $\mathcal{I}$ is the set of finite or scattered sets, we respectively say *finite-error* or *scattered-error*.

How well can a predictor do at accurately guessing values of f? If $X = Y = \mathbb{R}$ and only continuous functions are under consideration, then one can always guess $f(x)$ correctly by taking $\lim_{t \to x} f(t)$. If one allows arbitrary functions, then there is no hope of having a predictor S that always guesses correctly. In fact, we can make any predictor wrong on any scattered set; scattered sets are the topological analog of well-founded sets, and the proof of the following theorem is a straightforward diagonalization by induction.

THEOREM 1.1 ([HT09]). *Suppose $|Y| \geq 2$, S is a predictor, and $A \subseteq X$ is scattered. Then there is a function $f \in {}^X Y$ such that $A \subseteq Sf \triangle f$.*

Scattered sets tend to be small. For instance, in the usual topology on $\mathbb{R}$, they are exactly the countable G_δs [DG76]; in the upward topology on any ordinal, they are exactly the finite sets. So it is surprising that scattered-error predictors exist, provided X is T_0. The method of prediction introduced in [HT08] and generalized to topological spaces in [HT09] is the *μ-predictor*, which we now consider.

1.3. The μ-predictor

Definition 1.2. Fix a well-ordering $\preceq$ of ${}^XY$. For $f \in {}^XY$ and $x \in X$, let $\langle f\rangle_x$ be the $\preceq$-least element of $[f]_x$. The *μ-predictor* $M_\preceq$ (or simply M when $\preceq$ is understood) is defined by

$$Mf(x) = \langle f\rangle_x(x).$$

If one thinks of $f \prec g$ as meaning f is "simpler" than g, then one can see the μ-predictor as a formalization of Occam's razor: at each point, M guesses according to the $\preceq$-least ("simplest") function consistent with the information available.

Theorem 1.3. *If X is T_0 and $\preceq$ is any well-ordering of ${}^XY$, then M is a scattered-error predictor.*

Corollary 1.4. *If X is T_0, there exists a scattered-error predictor for ${}^XY$.*

We prove Theorem 1.3 at the end of this section, as it has not yet been published, but it is very similar to Theorem 2.4 of [HT09]; that result uses a slight modification of the μ-predictor to allow for a more economical proof (namely, it uses the μ^*-predictor, which ignores finite differences, and which we visit in Section 5).

Our main result is the following; it shows that when X is T_0, every scattered-error predictor is a special case of the μ-predictor.

Theorem 1.5. *Suppose X is T_0 and S is a scattered-error predictor for ${}^XY$. Then $S = M_\preceq$ for some well-ordering $\preceq$ of ${}^XY$.*

We actually prove something slightly stronger that involves a generalization of a couple of the above concepts to allow finer control over the information available to predictors; this generality is needed in Sections 4 and 5. A *notion of indistinguishability* $\equiv$ assigns to each $x \in X$ an equivalence relation $\equiv_x$ on ${}^XY$. Most notably, the relations $\approx_x$ above give a notion of indistinguishability $\approx$, which we take as our default if no other notion is specified. A function $S : {}^XY \to {}^XY$ *respects* $\equiv$ if (1.1) holds with $\equiv$ in place of $\approx$, and we call S a *predictor under* $\equiv$. In any notation defined in terms of $\approx$, we add a superscript to indicate the use of $\equiv$ in place of $\approx$; for example, $[f]_x^\equiv$ is the equivalence class of f under $\equiv_x$, and $M_\preceq^\equiv f(x) = \langle f\rangle_x^\equiv(x)$. Naturally, we say that $\equiv$ *refines* (*coarsens*) $\equiv'$ if $\equiv_x$ refines (coarsens) $\equiv'_x$ for each $x \in X$. Note that if $\equiv$ refines $\equiv'$, then any predictor under $\equiv'$ is also a predictor under $\equiv$. We can now state the stronger version of Theorem 1.5.

Theorem 1.6. *Suppose X is T_0 and S is a scattered-error predictor for ${}^XY$. Then there exists a well-ordering $\preceq$ of ${}^XY$ such that for any notion of indistinguishability $\equiv$ that coarsens $\approx$ and which S respects, $S = M_\preceq^\equiv$.*

In addition to characterizing the scattered-error predictors for T_0 spaces, the above results suggest a certain naturality to the μ-predictor. They also give some progress toward determining the strength of Corollary 1.4; in particular, does it imply AC over ZF? Our proofs of Theorems 1.5 and 1.6 are carried out in ZFC, but we examine what *can* be done in ZF in Section 6.

The rest of the paper is organized as follows. We finish the introduction with the proof of Theorem 1.3. In Section 2, we take a closer look at scattered sets and derive some results about scattered-error predictors that are needed in Section 3, where we prove the main results. In Section 4, we give a variation of Theorem 1.5 in a context where "visibility" is specified by a binary relation on X rather than a topology. Section 5 considers two variations of the μ-predictor, obtained by intentionally ignoring certain information, and explains the way in which the main results apply to them. Section 6 revisits some of the material from within ZF. We finish in Section 7 with further questions.

Proof of Theorem 1.3. Take any $f \in {}^XY$ and let $W = Mf \triangle f$.

With $\leq$ as defined in Section 1.1, we claim that for $x, y \in W$, $x < y \Rightarrow \langle f\rangle_x \prec \langle f\rangle_y$. Suppose $x, y \in W$ with $x < y$. Let V be a neighborhood of y (and hence also of x) witnessing $f \approx_y \langle f\rangle_y$. Since $x \not\geq y$, x has a neighborhood U with $y \notin U$. Then f and $\langle f\rangle_y$ agree on $U \cap V$, witnessing $f \approx_x \langle f\rangle_y$, so $\langle f\rangle_x \preceq \langle f\rangle_y$. However, $\langle f\rangle_y(x) = f(x)$ (since $x \in V - \{y\}$) and $\langle f\rangle_x(x) \neq f(x)$ (since $x \in W$), so $\langle f\rangle_x \neq \langle f\rangle_y$. We now have $\langle f\rangle_x \prec \langle f\rangle_y$, establishing the claim.

From the claim, it follows that W is well-founded in $\leq$; otherwise, it would induce an infinite descending chain in $\preceq$.

Suppose for a contradiction that W is not scattered. Then, as X is T_0, W is not weakly scattered, so there exists a nonempty $W' \subseteq W$ with no weakly isolated points. Note that any nonempty intersection of an open set with W' is infinite with no weakly isolated points.

Let $x \in W'$ be such that $\langle f\rangle_x$ is $\preceq$-minimal among $\{\langle f\rangle_y \mid y \in W'\}$, and let V be a neighborhood of x witnessing $f \approx_x \langle f\rangle_x$. Let $W'' = W' \cap V - \{x\}$, which is infinite.

Take any $y \in W''$ and suppose $y \not\geq x$. Then y has a neighborhood U excluding x. So $f \approx_y \langle f\rangle_x$, since $\langle f\rangle_x$ and f agree on $V - \{x\} \supseteq V \cap U$. It follows that $\langle f\rangle_y \preceq \langle f\rangle_x$, and hence $\langle f\rangle_y = \langle f\rangle_x$ by the minimality of $\langle f\rangle_x$. Consequently, M guesses correctly at y, since $\langle f\rangle_x$ and f agree on $V - \{x\}$, a contradiction. So, for all $y \in W''$, $y > x$.

Let Z be the set of $\leq$-minimal elements of W''. Since $\leq$ is well-founded on W, every $y \in W''$ has a $z \in Z$ with $z \leq y$. Note that Z forms an antichain in $\leq$.

Take any neighborhood V' of x. We claim that $V' \cap Z$ is infinite. If not, let $V' \cap Z = \{z_1, \ldots, z_n\}$. For $1 \le i \le n$, we have $x < z_i$, so there is a neighborhood V_i of x that excludes z_i. Then the set $T = (\bigcap_i V_i) \cap V \cap V'$ is a neighborhood of x disjoint from Z. Since the complement of T is closed and hence upwards closed under $\le$, T is in fact disjoint from W''. So, $T \cap W' = \{x\}$, contradicting the fact that W' has no weakly isolated points. Therefore, every neighborhood of x has infinite intersection with Z.

For any $z \in Z$ and neighborhood U of z, U is a neighborhood of x (since $x < z$), so U intersects Z infinitely as shown above. Hence Z has no weakly isolated points.

Take $z_0 \in Z$ such that $\langle f\rangle_{z_0}$ is $\preceq$-minimal among $\{\langle f\rangle_z \mid z \in Z\}$. Let U_0 be a neighborhood of z_0 witnessing $f \approx_{z_0} \langle f\rangle_{z_0}$. Take any $z_1 \in Z \cap U_0 - \{z_0\}$. Since Z is an antichain in $\le$, $z_0 \not\le z_1$, so z_1 has a neighborhood U_1 excluding z_0. Then f and $\langle f\rangle_{z_0}$ agree on $U_0 \cap U_1$, so $f \approx_{z_1} \langle f\rangle_{z_0}$, yielding $\langle f\rangle_{z_1} \preceq \langle f\rangle_{z_0}$. It follows that $\langle f\rangle_{z_1} = \langle f\rangle_{z_0}$ by the minimality of $\langle f\rangle_{z_0}$. So, since f and $\langle f\rangle_{z_0}$ agree on $U_0 - \{z_0\}$, M guesses correctly at z_1, a contradiction. ■

2. Scattered sets

DEFINITION 2.1. For $A \subseteq X$, let

$$\lim A = \{x \in X \mid \text{every deleted neighborhood of } x \text{ intersects } A\}.$$

Define $A^\bullet = A \cap \lim A$, and define $A^{(\alpha)}$ for ordinals α by

$$\begin{aligned} A^{(0)} &= A, \\ A^{(\alpha+1)} &= (A^{(\alpha)})^\bullet, \\ A^{(\lambda)} &= \bigcap_{\alpha<\lambda} A^{(\alpha)} \quad (\lambda \text{ a limit ordinal}). \end{aligned}$$

The *rank* of A is the least ordinal $\rho(A)$ such that $A^{(\rho(A)+1)} = A^{(\rho(A))}$; we call $A^{(\rho(A))}$ the *kernel* of A.

This is very similar to Cantor–Bendixson derivatives and rank, except that we have $A^\bullet = A \cap \lim A$, while the Cantor–Bendixson derivative of A is $\lim A$.

PROPOSITION 2.2. *A set is scattered iff its kernel is* $\emptyset$.

PROPOSITION 2.3. *In the downward (resp. upward) topology on a partial order, the scattered sets are the well-founded (resp. co-well-founded) sets.*

PROPOSITION 2.4. *If sets* $U_i \subseteq X$, $i \in I$, *are open and* $\Sigma \subseteq \bigcup_i U_i$, *then* Σ *is (weakly) scattered iff* $\Sigma \cap U_i$ *is (weakly) scattered for each* $i \in I$.

PROPOSITION 2.5. *The family* $\mathcal{I}$ *of weakly scattered subsets of* X *forms an ideal, i.e.,* $\emptyset \in \mathcal{I}$, $(A \in \mathcal{I}\ \&\ B \subseteq A) \Rightarrow B \in \mathcal{I}$, *and* $A, B \in \mathcal{I} \Rightarrow A \cup B \in \mathcal{I}$.

(In T_0 spaces, this ideal is the same as the family of scattered sets. In non-T_0 spaces, the latter is not an ideal: If x_0 and x_1 witness that X is not T_0, then $\{x_0\}$ and $\{x_1\}$ are scattered but $\{x_0, x_1\}$ is not.)

We define the relation $=^\dagger$ on ${}^XY$ by $f =^\dagger g$ iff $f \triangle g$ is weakly scattered. In light of Proposition 2.5, this is an equivalence relation, and we use $[f]^\dagger$ to denote the equivalence class of f under $=^\dagger$.

LEMMA 2.6. *If X is T_0, $\Sigma \subseteq X$ is scattered and $x \in X$, then x has a neighborhood V such that $V \cap \Sigma \cap \overline{x} - \{x\} = \emptyset$.*

Proof. Let $\Sigma' = \Sigma \cap \overline{x} - \{x\}$. If $\Sigma' = \emptyset$, then we can let V be any neighborhood of x. Otherwise, since Σ' is scattered, there exists $y \in \Sigma'$ with neighborhood W such that $W \cap \Sigma' = \{y\}$. Since $x \leq y$, every neighborhood of y contains x; in particular, W is a neighborhood of x, and since X is T_0 and $x \neq y$, x has a neighborhood U such that $y \notin U$. Let $V = W \cap U$. Then V is a neighborhood of x disjoint from Σ', so $V \cap \Sigma \cap \overline{x} - \{x\} = \emptyset$. ∎

2.1. Fixed points of scattered-error predictors. *Throughout this subsection, we assume X is T_0.*

PROPOSITION 2.7. *If U is open and $f|U = f'|U$, then $Sf|U = Sf'|U$ for any predictor S.*

LEMMA 2.8. *Suppose $x \in X$ and $f, f' \in {}^XY$ are such that $f \triangle f' \subseteq \{x\}$. Then for any scattered-error predictor S, x has a neighborhood V such that $(Sf \triangle Sf') \cap V = \emptyset$.*

Proof. Let $\Sigma = Sf \triangle Sf'$, which is scattered since $Sf =^\dagger f =^\dagger f' =^\dagger Sf'$. By Lemma 2.6, let V be a neighborhood of x such that $V \cap \Sigma \cap \overline{x} - \{x\} = \emptyset$. We claim that $\Sigma \cap V = \emptyset$; for this, it suffices to show that $\Sigma \subseteq \overline{x} - \{x\}$. If $y \notin \overline{x}$, then y has a neighborhood U with $x \notin U$; U witnesses $f \approx_y f'$, so $Sf(y) = Sf'(y)$ and hence $y \notin \Sigma$. Also, $f \approx_x f'$, so $x \notin \Sigma$. This establishes the claim and we now have $(Sf \triangle Sf') \cap V = \Sigma \cap V = \emptyset$. ∎

LEMMA 2.9. *If S is a scattered-error predictor, then every equivalence class of $=^\dagger$ contains exactly one fixed point of S.*

Proof. Let $h \in {}^XY$. We first show that S has at most one fixed point in $[h]^\dagger$. Suppose $f, f' \in [h]^\dagger$ are distinct fixed points. Then $f \triangle f'$ is nonempty but scattered. Let $x \in f \triangle f'$ with neighborhood V be such that $(f \triangle f') \cap V = \{x\}$. Then $f \approx_x f'$, hence $Sf(x) = Sf'(x)$; since $Sf = f$ and $Sf' = f'$, we then have $f(x) = f'(x)$, a contradiction. So S has at most one fixed point in $[h]^\dagger$.

It remains to be shown that a fixed point of S in $[h]^\dagger$ exists. Let

$$\mathcal{A} = \{U \subseteq X \mid U \text{ is open and } \exists f \in [h]^\dagger\ Sf|U = f|U\}.$$

We first show that no proper subset of X is a maximal element of $\mathcal{A}$, and then show that $\mathcal{A}$ is closed under arbitrary unions.

Take any $U \in \mathcal{A}$, and let $f \in [h]^\dagger$ be such that $Sf|U = f|U$. Assume $Sf \bigtriangleup f \neq \emptyset$ (otherwise, f is a fixed point and we are done). Since $Sf \bigtriangleup f$ is scattered, there exists $x \in Sf \bigtriangleup f$ with neighborhood V such that $(Sf \bigtriangleup f) \cap V = \{x\}$. Let $f' = f[Sf(x)/x]$. By Lemma 2.8, shrinking V if necessary, we can assume without loss of generality that $(Sf' \bigtriangleup Sf) \cap V = \emptyset$. For $z \in V - \{x\}$, we have $Sf'(z) = Sf(z) = f(z) = f'(z)$, and we have $Sf'(x) = Sf(x) = f'(x)$; this yields $Sf'|V = f'|V$.

For $z \in U$, Proposition 2.7 gives us $Sf'(z) = Sf(z) = f(z) = f'(z)$, hence $Sf'|U = f'|U$. We now have $Sf'|(U \cup V) = f'|(U \cup V)$, so f' witnesses that $U \cup V \in \mathcal{A}$. Note that U is a proper subset of $U \cup V$, since $x \in V - U$. So, no proper subset of X is maximal in $\mathcal{A}$.

We now show that $\mathcal{A}$ is closed under arbitrary unions. Suppose we have sets $U_i \in \mathcal{A}$ with associated functions $f_i \in [h]^\dagger$, for i in some index set I. We first show that the partial functions $f_i|U_i$ are compatible. If not, then there are $j, k \in I$ such that, letting $U' = U_j \cap U_k$, the set $\Sigma = (f_j \bigtriangleup f_k) \cap U'$ is nonempty. Since Σ is scattered, there exists $x \in \Sigma$ with neighborhood V such that $\Sigma \cap V = \{x\}$. Then $f_j \approx_x f_k$, so $Sf_j(x) = Sf_k(x)$, hence $f_j(x) = Sf_j(x) = Sf_k(x) = f_k(x)$, a contradiction.

Let $U = \bigcup_{i \in I} U_i$. Since the partial functions $f_i|U_i$ are compatible, their union is a function $f : U \to Y$. Extend f to a function $f : X \to Y$ by letting $f|(X - U) = h|(X - U)$. Take any $x \in U$, and let $i \in I$ be such that $x \in U_i$; noting that $f \approx_x f_i$, we have $Sf(x) = Sf_i(x) = f_i(x) = f(x)$. It follows that $Sf|U = f|U$, so f witnesses that $U \in \mathcal{A}$, provided $f =^\dagger h$, which follows from Proposition 2.4. This establishes the claim that $\mathcal{A}$ is closed under arbitrary unions.

Since $\mathcal{A}$ is closed under arbitrary unions (including the empty union, so $\emptyset \in \mathcal{A}$), and no proper subset of X is maximal in $\mathcal{A}$, it follows that $X \in \mathcal{A}$. The f witnessing $X \in \mathcal{A}$ is a fixed point of S. ∎

LEMMA 2.10. *Given any scattered-error predictor S (for ${}^XY$), any $f \in {}^XY$, and any $D \subseteq X$, there exists $f' \in {}^XY$ such that $f'|D = f|D$ and $Sf' \bigtriangleup f' \subseteq D$.*

The idea in the following proof is that fixing $f|D$ induces a scattered-error predictor for ${}^{(X-D)}Y$ to which we can apply Lemma 2.9.

Proof. Let $X_0 = X - D$ with the subspace topology. For any $h \in {}^{X_0}Y$, define $\hat{h} \in {}^XY$ by $\hat{h}|X_0 = h$, $\hat{h}|D = f|D$. Define the predictor S_0 for ${}^{X_0}Y$ by $S_0 g = (S\hat{g})|X_0$. It follows from the fact that S is a scattered-error predictor (for ${}^XY$) that S_0 is a scattered-error predictor (for ${}^{X_0}Y$), so by the previous lemma, there is an $h \in {}^{X_0}Y$ such that $h =^\dagger f|X_0$ and $S_0 h = h$. Then $S\hat{h}|X_0 = S_0 h = h = \hat{h}|X_0$, so $S\hat{h} \bigtriangleup \hat{h} \subseteq D$, and $\hat{h}|D = f|D$. We let $f' = \hat{h}$. ∎

3. Universality of the μ-predictor. *Throughout this section, we assume X is T_0.*

Fix a scattered-error predictor S, and let $\preceq$ be any well-ordering of ${}^XY$ such that $\rho(Sf \triangle f) < \rho(Sf' \triangle f') \Rightarrow f \prec f'$. We will show that the resulting $M_\preceq$ coincides with S (and, more generally, $M_\preceq^\equiv = S$ for appropriate $\equiv$). To get some intuition for how this will work, if we have $S\langle f\rangle_x(x) = \langle f\rangle_x(x)$, then it will follow that $Sf(x) = S\langle f\rangle_x(x) = \langle f\rangle_x(x) = Mf(x)$, which we want. In order to favor functions g where Sg and g agree at x, but without making specific reference to x (since we have one ordering $\preceq$ that is used at all points), we simply favor functions g where Sg and g agree often. In our case, the appropriate way to say that Sg and g agree often is to say that $\rho(Sg \triangle g)$ is small. By placing functions g with small values of $\rho(Sg \triangle g)$ early in the ordering, we will tend to get $S\langle f\rangle_x(x) = \langle f\rangle_x(x)$. That this is not just a tendency, but *always* happens, is worked out in the details that follow.

Suppose $\equiv$ is a notion of indistinguishability that coarsens $\approx$ but which is still respected by S (in the sense that $f \equiv_x g \Rightarrow Sf(x) = Sg(x)$).

Lemma 3.1. *Suppose $f \in {}^XY$, $x \in X$, and $g = \langle f\rangle_x^\equiv$. Then for any neighborhood V of x, $\rho((Sg \triangle g) \cap V - \{x\}) = \rho(Sg \triangle g)$.*

Proof. It is immediate that $\rho((Sg\triangle g)\cap V - \{x\}) \le \rho(Sg\triangle g)$. Suppose for a contradiction that $\rho((Sg\triangle g)\cap V - \{x\}) < \rho(Sg\triangle g)$. Let $g' = g[Sg(x)/x]$. By Lemma 2.8, let V' be a neighborhood of x such that

$$(Sg \triangle Sg') \cap V' = \emptyset. \tag{3.1}$$

Without loss of generality, $V' \subseteq V$. By Lemma 2.10, there is a $g'' \in {}^XY$ such that $g'|V' = g''|V'$ and $Sg'' \triangle g'' \subseteq V'$. Note that $Sg|V' = Sg'|V' = Sg''|V'$ by (3.1) and Proposition 2.7.

We claim that $Sg'' \triangle g'' \subseteq (Sg \triangle g) \cap V - \{x\}$. Take any $z \in Sg'' \triangle g''$. Then $z \in V' \subseteq V$, since $Sg'' \triangle g'' \subseteq V'$. Note that $g \approx_x g' \approx_x g''$ (the former because $g\triangle g' = \{x\}$, the latter because $g'|V' = g''|V'$), so $Sg''(x) = Sg(x) = g'(x) = g''(x)$, hence $x \notin Sg''\triangle g''$, so $z \ne x$. Also, $Sg(z) = Sg''(z) \ne g''(z) = g'(z) = g(z)$, so $z \in Sg\triangle g$. We now have $z \in (Sg\triangle g)\cap V - \{x\}$, establishing the claim.

It follows that $\rho(Sg'' \triangle g'') \le \rho((Sg \triangle g) \cap V - \{x\}) < \rho(Sg \triangle g)$, so $g'' \prec g$. Note, however, that $g'' \approx_x g \equiv_x f$, so $g'' \equiv_x f$, hence g is not the $\preceq$-least element of $[f]_x^\equiv$, a contradiction. ∎

Lemma 3.2. *Let Σ be a scattered set and suppose that $x \in X$ is such that $\rho(\Sigma \cap V - \{x\}) = \rho(\Sigma)$ for every neighborhood V of x. Then $x \notin \Sigma$.*

Proof. Let $\sigma = \rho(\Sigma)$. Let γ be minimal such that $x \notin \Sigma^{(\gamma)}$. Note that $\gamma \le \sigma$ since $\Sigma^{(\sigma)} = \emptyset$. Note also that γ cannot be a limit ordinal (since, for any limit ordinal λ, any point absent from $\Sigma^{(\lambda)}$ is already absent from $\Sigma^{(\alpha)}$ for some $\alpha < \lambda$).

Suppose for a contradiction $x \in \Sigma$. Then $\gamma \neq 0$, so $\gamma = \beta + 1$ for some β. Then $x \in \Sigma^{(\beta)}$ and x has a neighborhood V such that $\Sigma^{(\beta)} \cap V - \{x\} = \emptyset$. Hence $(\Sigma \cap V - \{x\})^{(\beta)} = \emptyset$, so $\rho(\Sigma \cap V - \{x\}) \leq \beta < \sigma = \rho(\Sigma \cap V - \{x\})$, a contradiction. Therefore, $x \notin \Sigma$. ∎

LEMMA 3.3. *Suppose* $f \in {}^X Y$, $x \in X$, *and* $g = \langle f \rangle_x^{\equiv}$. *Then* $Sg(x) = g(x)$.

Proof. Let $\Sigma = Sg \triangle g$, a scattered set. By Lemma 3.1, for any neighborhood V of x, $\rho(\Sigma \cap V - \{x\}) = \rho(\Sigma)$. By Lemma 3.2, $x \notin \Sigma$, so $Sg(x) = g(x)$. ∎

Proof of Theorem 1.6. With S, $\preceq$, and $\equiv$ as above, take any $f \in {}^X Y$ and $x \in X$. Let $g = \langle f \rangle_x^{\equiv}$. By the previous lemma, $Sg(x) = g(x)$. Since $g \equiv_x f$, we have $Sg(x) = Sf(x)$. Then $M_{\preceq}^{\equiv} f(x) = g(x) = Sg(x) = Sf(x)$. Therefore, $S = M_{\preceq}^{\equiv}$. ∎

Theorem 1.5 follows as the special case where $\equiv$ is $\approx$.

4. Visibility relations. Rather than using a topology on X to give a notion of indistinguishability, we can use a binary relation in the following way. Let V be an irreflexive binary relation on X; the intended meaning of xVy is that x *sees* y, in the sense that the value of $f(y)$ is available when trying to guess $f(x)$, and we accordingly call V a *visibility relation.* The common metaphor here is hats: we imagine that X is a set of agents who have hats placed on their heads (with Y being the set of hat colors), V specifies who can see which hats, and the agents must try to guess the colors of their own hats from the hats they can see. Letting $V(x)$ denote the set $\{y \in X \mid xVy\}$, we define the notion of indistinguishability $\sim$ by $f \sim_x g$ iff $(f \triangle g) \cap V(x) = \emptyset$ (informally: x cannot see any difference between f and g). In the context of visibility relations, $\sim$ is the default notion of indistinguishability; in particular, a *predictor for* V must now respect $\sim$ rather than $\approx$.

An important observation is that if V is a transitive visibility relation on X (that is, a strict partial order of X) and we put the upward topology on X, then $\sim$ and $\approx$ coincide. In short, transitive visibility is a special case of the topological context.

To speak of sets being scattered, we need to have a topology in mind. In the cases we examine, we will be using the upward topology induced by a certain partial order. So, recalling that the scattered sets in the upward topology on a partial order are the co-well-founded sets, the role played by scattered sets in previous sections is played by co-well-founded sets below.

What we are able to show is that when V is acyclic, "good" predictors (when they exist at all) are all special cases of the μ-predictor.

THEOREM 4.1. *Let V be an acyclic visibility relation on X, let V^+ denote its transitive closure, and suppose that S is a predictor for V such that $Sf \triangle f$ is co-well-founded in V^+ for all $f \in {}^XY$. Then $S = M_{\preceq}$ for some well-ordering $\preceq$ of ${}^XY$. (Now, of course, $M_{\preceq}$ refers to $M_{\preceq}^{\sim}$, not $M_{\preceq}^{\approx}$.)*

Proof. As V is acyclic, V^+ is a strict partial order of X. Consider X as a topological space under the upward topology induced by V^+, and let $\approx$ be the resulting notion of indistinguishability. Note that $\approx$ refines $\sim$, so S respects $\approx$. Also, as noted above, the scattered sets coincide with the sets co-well-founded in V^+. So, we can consider S as a scattered-error predictor under $\approx$. Applying Theorem 1.6, let $\preceq$ be a well-ordering of ${}^XY$ such that $S = M_{\preceq}^{\equiv}$ for any $\equiv$ that coarsens $\approx$ and which S respects. In particular, this applies when $\equiv$ is $\sim$, so $S = M_{\preceq}^{\sim}$. ∎

A case of particular interest is finite-error predictors. The question of which relations V admit a finite-error predictor is an ongoing one; specifically, we would like to know whether or not the following are equivalent for $|Y| \geq 2$:

(i) V admits a finite-error predictor;
(ii) there is no sequence of distinct $x_0, x_1, \ldots \in X$ such that $\neg x_i V x_j$ for $i \leq j$.

The direction (i)⇒(ii) always holds, and (ii)⇒(i) is known to hold when X is countable or V is transitive [Har11]. Also, if (ii)⇒(i) holds for acyclic V, then it holds for all V (since intersecting V with a well-ordering of X makes V acyclic while preserving (ii)). The following corollary tells us that, in the acyclic case, we can restrict our attention to instances of the μ-predictor when seeking a finite-error predictor.

COROLLARY 4.2. *Suppose V is an acyclic visibility relation on X and that S is a finite-error predictor for V. Then $S = M_{\preceq}$ for some well-ordering $\preceq$ of ${}^XY$.*

Proof. Finite sets are necessarily co-well-founded in any partial order, so Theorem 4.1 applies. ∎

There is no hope of extending Theorem 4.1 and Corollary 4.2 to visibility relations that contain cycles (except in degenerate cases where there are no such S to begin with, or $|Y| \leq 1$), as the following simple theorem shows. Say that two predictors S and S' are *almost the same* if $Sf \triangle S'f$ is finite for all $f \in {}^XY$. Note that if S and S' are almost the same and $\mathcal{I}$ is a nonprincipal ideal, then S is $\mathcal{I}$-error iff S' is $\mathcal{I}$-error.

THEOREM 4.3. *Suppose V has a cycle, S is a predictor for V, and $|Y| \geq 2$. Then there exists a predictor S' that is almost the same as S and is not a special case of the μ-predictor.*

Proof. We use the cycle to construct S' in a way that guarantees at least one error. Such a predictor cannot be a special case of the μ-predictor, because there is always at least one function that makes the μ-predictor correct everywhere: for any well-ordering $\preceq$ of ${}^XY$, if f_0 is the least function in the ordering, $M_{\preceq} f_0 = f_0$.

Let $x_0 V x_1 V \cdots V x_{k-1} V x_0$ be a cycle of V. Let $d : Y \to Y$ be such that $d(y) \neq y$ for all $y \in Y$. For $f \in {}^XY$, we define

$$S'f(x) = \begin{cases} f(x_{i+1}) & \text{if } x = x_i,\ i < k-1, \\ d(f(x_0)) & \text{if } x = x_{k-1}, \\ Sf(x) & \text{otherwise.} \end{cases}$$

Informally, in S', all agents in the cycle other than x_{k-1} assume their hat color is the same as the color of the next agent in the cycle, while x_{k-1} assumes it is not; everywhere else, S' agrees with S (so S' is almost the same as S). This guarantees at least one error: if S' were correct at every point in the cycle, we would have $f(x_0) = f(x_1) = \cdots = f(x_{k-1}) = d(f(x_0)) \neq f(x_0)$, a contradiction. ∎

5. Variations on the μ-predictor. We are also interested in modified versions of the μ-predictor. For example, we can form the μ^*-predictor, which is like the μ-predictor but ignores finite differences (that is, it respects $=^*$). One virtue of the μ^*-predictor is that while the proof of Theorem 1.3 is about one page, the proof of the analogous result for the μ^*-predictor is 11 lines [HT09]; that gives the μ^*-predictor, perhaps, a greater claim to being the "right" approach. Another virtue of the μ^*-predictor is that its willingness to overlook certain minor differences makes it work in some contexts where the μ-predictor can fail. For example, if one lets V be the complement of the identity relation on a set X, then the μ^*-predictor will always be finite-error, but the μ-predictor will typically not be; also, as noted below, the μ^*-predictor is weakly scattered-error even in non-T_0 spaces.

Taking this idea further, we can consider the $\mu^\dagger$-predictor, which ignores weakly scattered sets of differences. (This only makes sense in the topological context. Though we can make sense of the μ^*-predictor when working with visibility relations, we only consider the topological case below.)

Formally, under a given notion of indistinguishability $\approx$, let $\approx^*$ be the finest coarsening of $\approx$ in which each $\approx^*_x$ respects $=^*$; define $\approx^\dagger$ similarly. For a given well-ordering $\preceq$ of ${}^XY$, the μ^*-predictor under $\approx$ is $M^{\approx^*}_{\preceq}$, while the $\mu^\dagger$-predictor under $\approx$ is $M^{\approx^\dagger}_{\preceq}$.

Given a scattered-error predictor S that respects $=^*$ (resp. $=^\dagger$), we already know (provided X is T_0) that S must be a special case of the μ-predictor. By Theorem 1.6, we can also say that S must be a special case of the μ^*-predictor (resp. the $\mu^\dagger$-predictor).

Much else of what we already know about the μ-predictor also carries immediately over to the μ^*- and $\mu^\dagger$-predictors. As detailed below, the μ^*- and $\mu^\dagger$-predictors can be obtained as special cases of the μ-predictor under a finer topology. Our only concern is that, when we refine the topology, we might introduce new weakly scattered sets, so that while the μ^*- or $\mu^\dagger$-predictor is weakly scattered-error with respect to the finer topology, perhaps it is not weakly scattered-error with respect to the original topology. We show below that, for the refinements under consideration, no new weakly scattered sets are introduced, putting the concern to rest.

Definition 5.1. Given a topology $\mathcal{U}$ on X, let $\mathrm{WS}(\mathcal{U})$ denote the ideal of sets that are weakly scattered with respect to $\mathcal{U}$, let $\mathcal{U}^*$ be the coarsest refinement of $\mathcal{U}$ containing all cofinite sets (equivalently, the coarsest T_1 refinement of $\mathcal{U}$), and let $\mathcal{U}^\dagger$ be the coarsest refinement of $\mathcal{U}$ containing the complements of sets in $\mathrm{WS}(\mathcal{U})$.

Observe that the μ^*-predictor, under $\mathcal{U}$, is realized as the μ-predictor under $\mathcal{U}^*$; likewise for the $\mu^\dagger$-predictor and $\mathcal{U}^\dagger$. Note that $\mathcal{U}^*$ and $\mathcal{U}^\dagger$ are always T_1, even if $\mathcal{U}$ is not T_0 (this, once the theorem below is proved, shows that the μ^*-predictor and the $\mu^\dagger$-predictor are weakly scattered-error in any space).

Proposition 5.2. $\mathcal{U} \subseteq \mathcal{U}^* \subseteq \mathcal{U}^\dagger = \{U - K \mid U \in \mathcal{U} \;\&\; K \in \mathrm{WS}(\mathcal{U})\}$.

Theorem 5.3. $\mathrm{WS}(\mathcal{U}) = \mathrm{WS}(\mathcal{U}^*) = \mathrm{WS}(\mathcal{U}^\dagger)$.

Proof. By $\mathcal{U} \subseteq \mathcal{U}^* \subseteq \mathcal{U}^\dagger$, the inclusions $\mathrm{WS}(\mathcal{U}) \subseteq \mathrm{WS}(\mathcal{U}^*) \subseteq \mathrm{WS}(\mathcal{U}^\dagger)$ are trivial, so we must show $\mathrm{WS}(\mathcal{U}^\dagger) \subseteq \mathrm{WS}(\mathcal{U})$. Suppose $\Sigma \in \mathrm{WS}(\mathcal{U}^\dagger)$, and take any nonempty $\Sigma' \subseteq \Sigma$. We must show that Σ' has a point that is weakly isolated with respect to $\mathcal{U}$. Let $x \in \Sigma'$ with neighborhood $V \in \mathcal{U}^\dagger$ be such that $\Sigma' \cap V$ is finite. Then $V = U - K$ for some $U \in \mathcal{U}$ and $K \in \mathrm{WS}(\mathcal{U})$. Without loss of generality, $K \subseteq U$ (so $U = V \cup K$). If $\Sigma' \cap U$ is finite, we are done. Otherwise, $\Sigma' \cap K$ must be infinite; in particular, it is a nonempty subset of $K \in \mathrm{WS}(\mathcal{U})$, so there exists some $y \in \Sigma' \cap K$ with neighborhood $W \in \mathcal{U}$ such that $W \cap \Sigma' \cap K$ is finite. One can now verify that $U \cap W \in \mathcal{U}$ is a neighborhood of $y \in \Sigma'$ that weakly isolates y from Σ'. Therefore, $\Sigma \in \mathrm{WS}(\mathcal{U})$. ∎

6. In ZF. We would like to know whether Corollary 1.4 (quantified over all X and Y) implies AC over ZF. For this purpose, the main results are not immediately of any use, since they are theorems of ZFC. Though all of Section 2 can be carried out in ZF, we appeal to AC at the beginning of Section 3 when extending $\preceq$ to a well-ordering. What happens if we skip that step?

Suppose that, at the beginning of Section 3, we let $f \prec f' \Leftrightarrow \rho(Sf \triangle f) < \rho(Sf' \triangle f')$, without extending to a well-ordering. This would be a well-founded partial order of ${}^XY$; it would not be total (except in degenerate cases), but it would be total enough (when X is T_0, at least) to uniquely determine $M_{\preceq}^{\equiv}$: roughly speaking, if it did not uniquely determine $M_{\preceq}^{\equiv}$, then our proof of Theorem 1.5 would not work, since it uses an arbitrary extension of the above ordering to $\preceq$. A more rigorous justification follows.

Rather than letting $\langle f\rangle_x^{\equiv}$ be the $\preceq$-least element of $[f]_x^{\equiv}$, we now define $\langle f\rangle_x^{\equiv}$ to be the set of $\preceq$-minimal elements of $[f]_x^{\equiv}$. Fix some $y_0 \in Y$ (the case $Y = \emptyset$ is uninteresting). We define $M = M_{\preceq}^{\equiv}$ as follows: if every $g \in \langle f\rangle_x^{\equiv}$ agrees on the value of $g(x)$, we take this to be $Mf(x)$; otherwise, we let $Mf(x) = y_0$. (This latter case never occurs, but we cannot assume that yet.) In the statements of Lemmas 3.1 and 3.3 and the proof of Theorem 1.6, $g = \langle f\rangle_x^{\equiv}$ becomes $g \in \langle f\rangle_x^{\equiv}$. At the end of the proof of Lemma 3.1, the contradiction is now that g is not $\preceq$-minimal in $[f]_x^{\equiv}$, rather than "g is not the $\preceq$-least element of $[f]_x^{\equiv}$." With these modifications, we still reach the conclusion $S = M_{\preceq}^{\equiv}$ in the proof of Theorem 1.6. (Also note that, with the modified version of Lemma 3.3, every $g \in \langle f\rangle_x^{\equiv}$ agrees on the value of $g(x)$: for $g, g' \in \langle f\rangle_x^{\equiv}$, we have $g(x) = Sg(x) = Sg'(x) = g'(x)$; so, the y_0 case above never occurs.)

Therefore, while the existence of a scattered-error predictor for ${}^XY$ does not yield (in ZF) a well-ordering of ${}^XY$, it does yield a well-founded partial order $\preceq$ of ${}^XY$ under which the μ-predictor is well-defined.

7. Further questions. We have seen that, in the context of T_0 spaces, every scattered-error predictor is an instance of the μ-predictor, and that every instance of the μ-predictor is scattered-error. Transitive visibility relations can be seen as a special case of this. However, nontransitive visibility relations are not as well understood. What we have shown is that, for an acyclic visibility relation V, every *good* predictor (that is, one guaranteeing that the set of errors is co-well-founded in V^+) is an instance of the μ-predictor; it is not always the case, though, that every well-ordering $\preceq$ makes $M_{\preceq}$ a good predictor. First, some relations admit no good predictor at all (for example, with V the successor relation on $X = \omega$ and $|Y| \geq 2$, no predictor can guarantee even a single correct guess); second, even when good predictors exist, $M_{\preceq}$ will be good for some choices of $\preceq$, but typically not all when V is nontransitive. So, a few questions arise: Which visibility relations admit good predictors? When a visibility relation admits at least one good predictor, which well-orderings $\preceq$ make $M_\preceq$ a good predictor? Even if we cannot answer the latter question fully, can we at least find a way to construct $\preceq$ such that, if there is any good predictor at all, then $M_\preceq$ is good?

Currently, the only known technique for producing good predictors based on the μ-predictor for nontransitive visibility relations is to voluntarily coarsen the notion of indistinguishability to one that is more cooperative, without coarsening it too much. For example, given a nontransitive visibility relation V, we can often find a transitive $T \subseteq V$ that is "close" to V in some sense, and use the μ-predictor with T as our notion of visibility; see [Har11] for details. In that same paper, an example is given for which that approach cannot be made to work; specifically, a nontransitive V is constructed that holds some promise for admitting a finite-error predictor, but for which no transitive subrelation admits a finite-error predictor. Yet we know from Corollary 4.2 that if any finite-error predictor exists, it can be realized as a special case of the μ-predictor. This is some of our motivation for the above questions: in situations where restricting to a transitive subrelation is not an option, we would like a way of constructing orderings $\preceq$ that make the μ-predictor perform well even in the absence of transitivity.

In the case of visibility relations with a cycle, we saw in Theorem 4.3 how predictors can fail to be special cases of the μ-predictor. Nevertheless, can we identify the circumstances under which, given a predictor S, there exists an instance of the μ-predictor that is "as good" as S? For example, is it the case that if $\mathcal{I}$ is an ideal and $\equiv$ is a notion of indistinguishability that admits an $\mathcal{I}$-error predictor, then there is an instance of the μ-predictor that is $\mathcal{I}$-error?

Separately, as considered earlier: Does Corollary 1.4, quantified over all X and Y, imply AC over ZF?

Acknowledgements. The author would like to thank Alan Taylor for helpful comments on earlier versions of this paper.

References

[DG76] R. O. Davies and F. Galvin, Solution to Query 5, Real Anal. Exchange 2 (1976), 74–75.

[Har11] C. S. Hardin, *On transitive subrelations of binary relations*, J. Symbolic Logic 76 (2011), 1429–1440.

[HT08] C. S. Hardin and A. D. Taylor, *A peculiar connection between the Axiom of Choice and predicting the future*, Amer. Math. Monthly 115 (2008), 91–96.

[HT09] C. S. Hardin and A. D. Taylor, *Limit-like predictability for discontinuous functions*, Proc. Amer. Math. Soc. 137 (2009), 3123–3128.

[Mor90] J. C. Morgan II, *Point Set Theory*, Dekker, New York, 1990.

Christopher S. Hardin
1 New York Plaza, 33rd Floor
New York, NY 10004, U.S.A.
E-mail: chardin@janestreet.com

Received 20 December 2011;
in revised form 14 February 2013