A Game-Theoretic Analysis of the ESP Game*
PRELIMINARY DRAFT: COMMENTS WELCOME

Shaili Jain David C. Parkes
School of Engineering and Applied Science
Harvard University
{shailij, parkes}@eecs.harvard.edu

October 7, 2009

Abstract

In recent years, there has been a great deal of progress in "Games with a Purpose," interactive games that users play because they are fun, with the added benefit that they are doing useful work in the process. The ESP game, developed by von Ahn and Dabbish [7], is an example of such a game devised to label images on the web. Since labeling images is a hard problem for computer vision algorithms and can be tedious and time-consuming for humans, the ESP game provides humans with incentive to do useful work by being enjoyable to play. We present a simple game-theoretic model for the ESP game and characterize the equilibrium behavior of the model. We show that a simple change in the incentive structure can lead to different equilibrium structure. Our results suggest the possibility of formal incentive design in achieving desirable system-wide outcomes in this area of "human computation" in complementing existing considerations of robustness against cheating and human factors.

1 Introduction

The paradigm of human computation considers the possibility that networks of people can be leveraged in solving large-scale problems that are hard for computers to solve. Showcased by the early success of "Games with a Purpose" [6], human computation provides an example of the broader agenda of "peer production" which seeks to design and understand the problem of promoting large-scale collaborations by humans outside of the traditional framework of firms and price signals [1]. Examples of peer production systems include Wikipedia and YouTube.

Work by von Ahn and others has shown the tremendous power that networks of humans possess to solve *problems* while playing computer games [7, 10, 8, 9]. The ESP game is an example of such human computation; it is an interactive system that allows users to be paired to play games that label images on the web [7]. Users play the ESP game because it is an enjoyable game to play, with the added side-effect that they are doing useful work in the process. Subsequent work to the ESP game has included Peekaboom [10], a two-player game for locating objects within an image, Phetch [8], a two to four-player game for gathering useful descriptions for images on the web, and Verbosity [9], a two-player game for gathering common sense facts. Still in the spirit of human computation, Kearns et al. [5] run a number of behavioral experiments to see how fast distributed networks of humans can solve various graph problems, such as graph coloring. The

*A preliminary version of this paper appeared in the Proceedings of the 4th International Workshop on Internet and Network Economics (WINE 2008).

authors study how network topology and information constraints change the relative difficulty of various graph problems for these distributed networks of humans.

While there has been incredible progress in the area of human computation, there is still much more potential. For Games with a Purpose it seems especially appropriate to use game theory to better understand how to design incentives in order to achieve system-wide goals. For example, it appears anecdotally that during play of the ESP game that people coordinate on easy words and that the game is less effective in labeling less obvious, harder words.[1] For more general settings of human computation and peer production it may also be necessary to embrace the methods of behavioral economics, including such considerations as other-regarding preferences and altruism [3, 1].

This paper aims to study behavior in the ESP Game through a game-theoretic light. We propose a simple model of the game and consider two different models of payoffs, namely *match-early* preferences and *rare-words-first* preferences. Match-early preferences model the setting in which players wish to complete as many rounds as possible and receive the same score irrespective of the words on which they match. The match-early preferences model is meant to reflect the current method of assigning scores to outcomes in the ESP game. Here we show that *low effort* is a Bayesian-Nash equilibrium for all distributions on word frequencies, with players focusing attention on high-frequency words. Rare-words-first preferences model the setting in which players wish to match on infrequent words before frequent words, we suppose because of appropriately designed incentives, and the speed with which a match is achieved is only a secondary consideration. We show that under this preference model, there is a significant difference in the equilibrium structure in that we can no longer focus on ordinal preferences and must assume more structure on the valuation model. Moreover we show that under the rare-words-first preferences, we are able to identify, for additional structure on valuation models, an equilibrium behavior that shows a useful focusing on lower frequency words. This is obtained for a broader range of distributions than is possible under the match-early preference model under similar structural assumptions.

In related work, Hsu and colleagues [4, 2] developed a simple game called PhotoSlap, for determining content of images, based on the popular card game Snap. Hsu et al. have provided a game-theoretic analysis for PhotoSlap and are able to establish that the desired behavior from a system-wide perspective is a subgame perfect Nash equilibrium. To our knowledge, the work of Hsu and colleagues is the first application of game theory to human computation, however their model and analysis are specific to their game and cannot be applied to the ESP game. The ESP game appears to require a more intricate analysis.

2 The ESP Game

The ESP Game [7] is a two-player game for labeling images on the web. Labeling images has proven to be a hard problem for computer vision, yet it is something that humans can do easily. However, in order to label images, humans require some sort of *incentive* for this normally tedious task.

Players are randomly paired and each player is presented with the same image. Once the two players have entered a common word, this common word becomes the label for the image. *Players cannot communicate with each other while they are entering words for the image and once they agree on a common word, they only see the common word that they agreed upon.* Players are paired for a set of 15 images and each pair tries to "label" as many of the images as they can in 2.5 minutes. Players receive a fixed number of points after agreeing on a common word. In the set of 15 images, players get bonus points after agreeing on five images, ten images, and fifteen images in the same set. In the set of 15 images, players can pass on difficult images and they are revisited at the end of the set. The only words that are used from the input streams are

[1] We recently learned that in Google's version of the ESP game, they reward more "descriptive words" to counter this unfortunate default behavior. It is unclear to us how Google Image Labeler defines how descriptive a word is and how exactly they compute the reward for more descriptive words.

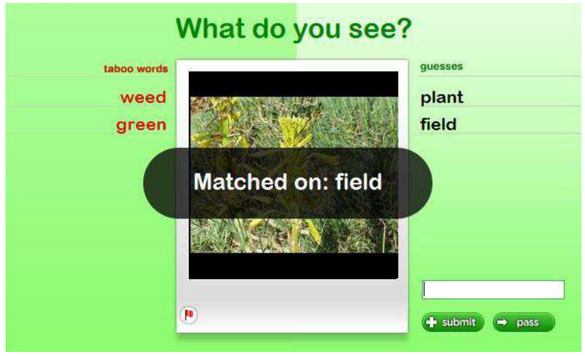

Figure 1: Sample play of ESP Game with Taboo Words shown on the left and player's input stream shown on the right.

the first agreed word. It is intuitive that words upon which players will agree are likely to be relevant to the image given that it is the image, and nothing else, around which the players can coordinate. Figure 1 shows the interface that players see when they play the ESP game and a sample input stream for a pair of players. The game includes a scoreboard, with the names of players with the highest scores, that is updated daily.

An interesting feature of the ESP game is the use of *Taboo Words*. Taboo words, as the name implies, are words that are displayed next to the image that players cannot use to gain points. From the system designer's perspective, Taboo words are words that have been entered sufficiently many times in previous plays of the image. Taboo words encourage players to enter different words so that the set of labels for an image can be extended. Another model parameter at the system designer's discretion is the *label threshold*. This determines how many times a word must be entered before it becomes a Taboo word.

von Ahn and Dabbish [7] put a number of methods in place to circumvent cheating in the ESP game. The fact that players cannot communicate with each other prevents collusion. In addition, IP addresses are recorded so that players cannot be paired with themselves. In order to prevent players from being paired with their friends who log in at the same time as them, the game server starts a new game every 30 seconds and when a new player logs in, the game server waits until the next 30-second boundary to pair the new player with another player. Once players agree on a label for a particular image, that label cannot be used for the rest of the set of 15 images, preventing cheating through the use of widespread "global" strategies. Another mechanism in place to prevent the use of global strategies, is measuring the average time it takes for players to agree. If this time substantially decreases, it may indicate the use of a large-scale unified strategy for cheating [7]. An example of such a unified strategy would be always entering "Obama" regardless of the image (such a strategy could be posted on a commonly accessed webpage). If a large-scale unified strategy is detected, the game server pairs each player with a bot (pre-recorded game play), to make cheating impossible.

From a game-theoretic perspective and system-design perspective, there are many interesting facets to the ESP game. For example, how does prior game play affect current game play? How does the use of Taboo words affect game play? How can we assign scores to outcomes to modify preferences on the kinds of outcomes achieved so as to promote desired behavior? This paper examines the latter question.

3 An ESP Model with Match-Early Preferences

We model the ESP game as a two-player game of imperfect information. We focus on modeling one of the 15 rounds, and thus the game associated with a specific image. We model the ESP game with each player sampling words from a universe of possible words associated with the picture, to which we associate a frequency ordering. Players can vary the effort level that relates to how likely they are to sample frequent words as opposed to infrequent words. Then players decide which order to play their sampled words in the game. In this model of *match-early* preferences, we instead capture the strategic behavior of having 15 rounds under a time constraint by providing a preference for matching in an earlier location rather than a later location.

Let $d > 0$ denote the *dictionary size* in our model, representing the number of words that each player will choose to think of for the image at hand. We model a universe of n words $U = \{w_1, w_2, ..., w_n\}$. These are the words that are in some way relevant to the image and represent the knowledge that the game designer is trying to learn. Each word has an associated frequency, where f_i denotes the frequency of word w_i. We assume that a player can rank the words sampled by frequency. The frequencies satisfy the property that $\sum_{i=1}^{n} f_i = 1$. We assume that the words in the universe are ordered according to decreasing frequency, that is $f_1 \geq f_2 \geq ... \geq f_n$. The frequency of word i can be considered the frequency with which the word would be mentioned if a very large population of humans were each asked to state a word related to the image. We assume that $1 < d < |U|$, since if $d = |U|$, players have complete information about their opponents and if $d = 1$, the game-theoretic analysis becomes trivial.

Even though this is a game without any communication between players, it is nevertheless useful to decompose the strategy of a player into two components which we associated with a *first stage*, i.e. choosing an effort level, and a *second stage*, i.e. choosing a permutation on a sampled dictionary.

In the first stage, a player chooses an *effort level*: $E = \{L, M, H\}$ for *low*, *medium* or *high*. The choice of effort level determines the set of words in the universe from which a player samples her dictionary. If a player chooses L in the first stage, the dictionary is sampled from the top $n_L > 0$ words (without replacement). That is, word i in the top n_L words is chosen first with probability $f_{i,L} = \frac{f_i}{\sum_{j=1}^{n_L} f_j}$. We use the notation that a player that chooses effort level L has universe U_L, where U_L is exactly the set of the highest n_L frequency words in U. In addition, let $\mathcal{D}_L$ denote the set of all possible dictionaries a player could obtain if they played L effort. If a player chooses effort M, the dictionary is sampled from the top $n_M > n_L > 0$ words, without replacement. That is, word i in the top n_M words is chosen with probability $f_{i,M} = \frac{f_i}{\sum_{j=1}^{n_M} f_j}$. Likewise, a player that chooses effort level M has universe U_M, where U_M is exactly the set of the highest n_M frequency words in U, and $\mathcal{D}_M$ denotes the set of all possible dictionaries a player could obtain if they played M effort. If a player chooses effort H, the dictionary is sampled from the top $n_H = n$ words (i.e., the entire universe), without replacement. That is word i in U is chosen with probability $f_{i,H} = f_i$. Note that we assume $d < n_L$, so a player has uncertainty about the dictionary sampled by the other player even in a low effort equilibrium and the game remains one of incomplete information.

Given a word $x \in U$, we let $f_e(x)$ represent the frequency of word x given that the player has chosen effort level e. In the game tree in Figure 2, this sample is modeled as a move by nature and can be considered to be the point at which a player learns her "type", namely her dictionary of words. Figure 2 represents the choices of a single player in the game, though both players are symmetric and the second player's game tree is identical. Note that n_L, n_M, n_H, and d are parameters of the model and that there is no cost associated with each of the first level actions. We establish that low effort is an equilibrium under match-early preferences even without introducing a cost, which would increase with effort and presumably increase the benefits of low effort. We leave introducing cost into this model for future work.

In the second stage, once each player privately learns her dictionary based on the effort level chosen, players choose a *permutation* on the words. This models the decision in the ESP game about the order in which a player should enter words. This order on a player's dictionary defines the second-stage action of each player and determines the *outcome* of the game. The outcome is defined by the first word that is in the ordered-list of both players and the location (where the location is defined to be the maximum value

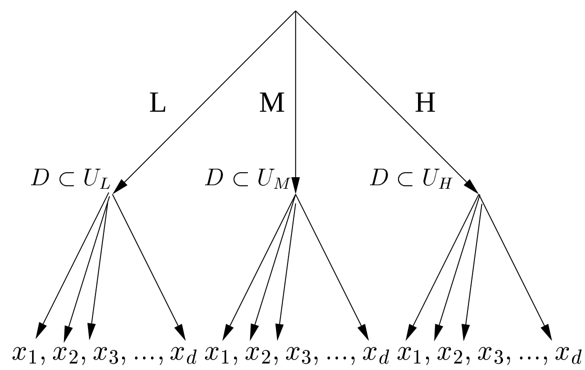

Figure 2: The game tree above represents the strategic problem of one player

of the two positions where the word occurs in each ordered-list) at which that occurs. In what follows, we refer to D_1 as the dictionary for player 1 and D_2 as the dictionary for player 2. The second stage strategy $s_1 \in S_1$ for player 1 defines a specific order $s_1(D_1)$ on D_1, for every possible dictionary. Given an effort level, which induces a distribution on sampled dictionaries, the second-stage strategy of a player defines a specific order in which words are played, for every possible dictionary. Likewise, player 2 has a second-stage strategy $s_2 \in S_2$ that defines an order on every possible dictionary. We restrict our attention to strategies that involve playing all words in the dictionary since any strategy that does not involve playing all words is weakly dominated by one that involves playing all words. Moreover, we will look for equilibrium in *consistent* strategies, which are strategies for a player that do not change the relative ordering of elements depending on the player's realized dictionary. In other words, a consistent strategy involves specifying a total ordering of elements on U_e and applying that total ordering to the realized dictionary.[2] A consistent strategy s specifies a total ordering on the set U_e: $w'_1 \succ w'_2 \succ \ldots \succ w'_{|U_e|}$, where w'_i is not necessarily the same as w_i. In fact, $w'_i = w_i$ for all i if and only if $s = s^{\downarrow}$, where $s^{\downarrow}$ is the strategy in which a player plays their words in order of decreasing frequency.

A complete strategy for the ESP game is a pair $\sigma_i = (e_i, s_i) \in E \times S_i = \Sigma_i$. This defines the play in both stages, with the second-level strategy s_i defining the order in which words in the dictionary are played for all possible dictionaries sampled under effort level e_i. We focus on pure strategies, because they are easier to justify than mixed strategies and because they exist in our game.

Definition 1. *We define a match as follows: Suppose player 1 outputs a list of words $x_1, x_2, ..., x_d$ and player 2 outputs a list of words $y_1, y_2, ..., y_d$. If there exists $1 \leq i, j \leq d$ such that $x_i = y_j$, then there was a match in this round of the game and we say that this match occurred in location $max(i, j)$. It is possible for two sequences to have more than one match, so we concern ourselves with the first match, that is the pair i, j that minimizes $max(i, j)$ such that $x_i = y_j$.*

Given this, an *outcome* is a pair $o = (w, l) \in (U \cup \phi) \times (\{1, ..., d\} \cup \phi)$ where (ϕ, ϕ) indicates there was no match and the (w, l) pair otherwise indicates that the *first* match occurred on word $w \in U$ in location $l \in \mathcal{L}$, where $\mathcal{L} = \{1, 2, \ldots, d\} \cup \phi$. Let $\mathcal{O}$ denote the set of possible outcomes. Since $s_1(D_1)$ and $s_2(D_2)$ specify orderings on given dictionaries, they induce an outcome: the location of the first match. Let *outcome function* $g(s_1(D_1), s_2(D_2)) \in \mathcal{O}$ denote this outcome, as defined by the rules of a first match. The location (if any) of the first match is denoted by $g_l(s_1(D_1), s_2(D_2)) \in \mathcal{L}$.

[2]Note that we do not restrict agents to only playing consistent strategies, but rather identify equilibria in which a player does not wish to deviate to an inconsistent strategy.

Each player wants to maximize her payoff. We model this by associating each player i with a *valuation function* $v_i(o)$ on outcome o, which induces a (weak) total preference ordering on outcomes. For match-early preferences, we require $(w_1, l_1) \equiv (w_2, l_1) \equiv ... \equiv (w_n, l_1) \succ (w_1, l_2) \equiv (w_2, l_2) \equiv ... \equiv (w_n, l_2) \succ ... \succ (w_1, l_d) \equiv (w_2, l_d) \equiv ... \equiv (w_n, l_d) \succ (\phi, \phi)$ for all players. With this, players prefer to match with their opponent as opposed to not matching, and players prefer to match in an earlier location rather than a later location. Players are agnostic as to *which* word is matched and care only about location.

Let $\Pr(D_i|e_i)$ denote the probability of dictionary D_i given effort level e_i. Often times we write this as $\Pr(D_i)$ and leave the effort level implicit. Given this, we now define the probability of first match in a particular location when player i knows her own type but has only probabilistic information on the dictionary of the other player.

Definition 2. *The probability of first match in location l_i given $s_1(D_1)$, s_2, and a distribution on dictionaries $\Pr(D_2)$, is $p(l_i, s_1(D_1), s_2) = \sum_{D_2} \Pr(D_2) I(g_l(s_1(D_1), s_2(D_2)) = l_i)$. Similarly, the probability of first match in location l_i on word w_j is $p(w_j, l_i, s_1(D_1), s_2) = \sum_{D_2} \Pr(D_2) I(g(s_1(D_1), s_2(D_2)) = (w_j, l_i))$. Often times we will abbreviate $p(l_i, s_1(D_1), s_2)$ as $p(l_i)$ and $p(w_j, l_i, s_1(D_1), s_2)$ as $p(w_j, l_i)$.*

Let $u_i(s_i(D_i), s_{-i}(D_{-i})) = v_i(g(s_1(D_1), s_2(D_2)))$ denote the utility to player i given realized dictionaries D_1 and D_2. Let $u_i(s_i(D_i), s_{-i}) = \sum_{D_{-i}} \Pr(D_{-i}) u_i(s_i(D_i), s_{-i}(D_{-i}))$ denote the expected (interim) utility to player i given dictionary D_i but with respect to a distribution on the possible dictionary of the other player, as induced by her effort level. It is also helpful to adopt $u_i(\sigma_i, \sigma_{-i}) = \sum_{D_1} \sum_{D_2} \Pr(D_1|e_1) \Pr(D_2|e_2) u_i(s_i(D_i), s_{-i}(D_{-i}))$ to denote the expected utility to player i before either dictionaries are sampled, given complete strategies $\sigma = (\sigma_1, \sigma_2)$.

In analyzing the equilibrium of the ESP game it will be helpful to isolate a restricted game, which is that induced for a fixed pair of first stage strategies (i.e., efforts) of each player. For a complete strategy profile (σ_1, σ_2) to be an equilibrium, it is necessary that neither player can usefully deviate to an alternate second-stage strategy. Of course this is not sufficient to establish an equilibrium of the full game, in that player might usefully deviate to an alternate effort, perhaps in combination with an alternate second stage strategy. To continue, consider the game induced by fixing top level effort levels (e_1, e_2) for the two players. This is a restricted game, that we refer to here as the *second stage game*, which is conditioned on effort e_1 and e_2. In this second stage game, each player knows her own dictionary but not the dictionary of the other player. Given this, we can define two useful equilibrium concepts:

Definition 3. *Second-stage strategy profile $s^* = (s_1^*, s_2^*)$ is an* ex post *Nash equilibrium of the second stage of the ESP game conditioned on effort levels e_1 and e_2, if for every D_1 and every D_2, we have:*

$$u_i(s_i^*(D_i), s_{-i}^*(D_{-i})) \geq u_i(s_i'(D_i), s_{-i}^*(D_{-i})),\ \forall s_i' \neq s_i^*, \quad \forall i \in \{1, 2\} \tag{1}$$

This states that as long as player 2 is adopting the equilibrium strategy then player 1 never regrets playing her sampled words in the order prescribed by strategy $s_1(D_1)$.

Definition 4. *Second-stage strategy profile $s^* = (s_1^*, s_2^*)$ is a strict Bayesian-Nash equilibrium of the second-stage of the ESP game conditioned on effort levels e_1 and e_2 if for both players $i \in \{1, 2\}$, for every D_i,*

$$u_i(s_i^*(D_i), s_{-i}^*) > u_i(s_i'(D_i), s_{-i}^*), \tag{2}$$

where the probability adopted in interim utility u_i for the distribution on the dictionary of player $-i$ is induced by the effort of that player in the first stage.

This states that player 1 strictly maximizes her expected utility, where the expectation is taken with respect to the possible dictionaries of player 2 and conditioned on the effort level of player 2. We also define a Bayesian-Nash equilibrium for the entire game.

Definition 5. *A strategy profile $\sigma^* = (\sigma_1^*, \sigma_2^*) \in \Sigma_1 \times \Sigma_2$ is a strict Bayesian-Nash equilibrium of the ESP game if for both players $i \in \{1, 2\}$, we have $u_i(\sigma_i^*, \sigma_{-i}^*) > u_i(\sigma_i', \sigma_{-i}^*)$.*

Since the effort level chosen by each player is not visible to the other player, there is no need for a subgame perfect refinement.

We will adopt an analysis approach that establishes a strict *ordinal* Bayesian-Nash equilibrium, in the sense that we identify strategies that are an equilibrium for all valuation functions consistent with match-early preferences. This is introduced in the next section, along with an associated property of stochastic dominance with respect to a distribution on outcomes in the game.

4 Effort Level of Players under Match-Early Preferences

In this section, we analyze the equilibrium behavior under match-early preferences. We show that playing *decreasing frequency* in conjunction with *low effort* is a Bayesian-Nash equilibrium for the ESP game. First we see that playing words in order of decreasing frequency is not an ex-post Nash equilibrium for the second stage game.

Lemma 1. *Suppose that players are playing the same effort level e and there are three words in the universe, $U_e = \{w_1, w_2, w_3\}$. The second-stage strategy profile $s = (s_1, s_2)$, where s_1 and s_2 are the strategies of playing words in order of decreasing frequency, is not an ex-post Nash equilibrium.*

Proof. Suppose player 2 samples dictionary $D_2 = \{w_2, w_3\}$ and player 1 samples dictionary $D_1 = \{w_1, w_2\}$. Strategy s_2 dictates player 2 will play w_2 followed by w_3. If player 1 deviates from s_1 and plays w_2 first followed by w_1, then player 1 will get higher utility. ■

Since playing words in order of decreasing frequency is not an ex-post Nash equilibrium, we focus instead on Bayesian-Nash equilibrium. The following definition of *stochastic dominance* will enable equilibrium analysis for every valuation function that satisfies MEP. In Lemmas 2 and 3, we show that this notion of stochastic dominance is both sufficient and necessary for utility maximization.

Definition 6. *Fixing effort levels e_1 and e_2, fixing the opponent's second-stage strategy s_2, and fixing dictionary D_1, we say that the second-level ordering $s_1(D_1)$ with probabilities of a first match $(p(l_1, s_1(D_1), s_2)$, $p(l_2, s_1(D_1), s_2)$, ..., $p(l_d, s_1(D_1), s_2))$ stochastically dominates the second-level ordering $s_1'(D_1)$ with probabilities of a first match $(p(l_1, s_1'(D_1), s_2)$, $p(l_2, s_1'(D_1), s_2)$, ..., $p(l_d, s_1'(D_1), s_2))$ if and only if $\sum_{a=1}^{k} p(l_a, s_1(D_1), s_2) \geq \sum_{a=1}^{k} p(l_a, s_1'(D_1), s_2)$ for every $1 \leq k \leq d$. We say that the stochastic dominance property is strict if there exists a k such that $1 \leq k \leq d$ and $\sum_{a=1}^{k} p(l_a, s_1(D_1), s_2) > \sum_{a=1}^{k} p(l_a, s_1'(D_1), s_2)$.*

Definition 6 defines the notion of stochastic dominance for an ordering $s_1(D_1)$. It should be noted that we say a second-level strategy s_1 stochastically dominates another second level strategy s_1' if and only if $s(D_1)$ stochastically dominates $s_1'(D_1)$ for all $D_1 \in \mathcal{D}_1$.

Lemma 2. *If the ordering $s_1(D_1)$ stochastically dominates the ordering $s_1'(D_1)$, for fixed opponent second-stage strategy s_2, then $u_1(s_1(D_1), s_2) \geq u_1(s_1'(D_1), s_2)$, for all valuations consistent with match-early preferences.*

Proof. Since $s_1(D_1)$ stochastically dominates $s_1'(D_1)$, we have for every $1 \leq k \leq d$, $\sum_{a=1}^{k} p(l_a, s_1(D_1), s_2) \geq \sum_{a=1}^{k} p(l_a, s_1'(D_1), s_2)$. Since $p(l_1, s_1(D_1), s_2) \geq p(l_1, s_1'(D_1), s_2)$, we have that $p(l_1, s_1(D_1), s_2)v(l_1) \geq p(l_1, s_1'(D_1), s_2)v(l_1)$. For $1 \leq k < d$, $p(l_{k+1}, s_1(D_1), s_2) \geq \sum_{a=1}^{k} p(l_a, s_1'(D_1), s_2) - \sum_{a=1}^{k} p(l_a, s_1(D_1), s_2) + p(l_{k+1}, s_1'(D_1), s_2)$. Consider

$$\sum_{a=1}^{k+1} p(l_a, s_1(D_1), s_2)v(l_a) = \sum_{a=1}^{k} p(l_a, s_1(D_1), s_2)v(l_a) + p(l_{k+1}, s_1(D_1), s_2)v(l_{k+1})$$

$$\geq \sum_{a=1}^{k} p(l_a, s_1(D_1), s_2)v(l_a) + (\sum_{a=1}^{k} p(l_a, s_1'(D_1), s_2) - \sum_{a=1}^{k} p(l_a, s_1(D_1), s_2) + p(l_{k+1}, s_1'(D_1), s_2))v(l_{k+1})$$

$$\geq \sum_{a=1}^{k} p(l_a, s_1(D_1), s_2)(v(l_a) - v(l_{k+1})) + \sum_{a=1}^{k} p(l_a, s_1'(D_1), s_2)v(l_{k+1}) + p(l_{k+1}, s_1'(D_1), s_2)v(l_{k+1})$$

$$\geq \sum_{a=1}^{k} p(l_a, s_1'(D_1), s_2)(v(l_a) - v(l_{k+1})) + \sum_{a=1}^{k} p(l_a, s_1'(D_1), s_2)v(l_{k+1}) + p(l_{k+1}, s_1'(D_1), s_2)v(l_{k+1})$$

$$= \sum_{a=1}^{k} p(l_a, s_1'(D_1), s_2)v(l_a) + p(l_{k+1}, s_1'(D_1), s_2))v(l_{k+1})$$

Therefore $u_1(s_1(D_1), s_2) \geq u_1(s_1'(D_1), s_2)$. ■

Our definition of stochastic dominance is not only sufficient, but it is necessary.

Lemma 3. *If $u_1(s_1(D_1), s_2) \geq u_1(s_1'(D_1), s_2)$ for all valuations that are consistent with match-early preferences, then ordering $s_1(D_1)$ must stochastically dominate ordering $s_1'(D_1)$, for fixed opponent second-stage strategy s_2.*

Proof. For the purpose of the proof adopt $p(l_a, s_1)$ as shorthand for $p(l_a, s_1(D_1), D_2)$, and also $p(l_{\leq a}, s_1) = \sum_{a'=1}^{a} p(l_{a'}, s_1)$, $p(l_{>a}, s_1) = \sum_{a'=a+1}^{d} p(l_{a'}, s_1)$. Suppose for contradiction that there is some k, $1 \leq k < d$ for which stochastic dominance is violated and $\sum_{a=1}^{k} p(l_a, s_1) < \sum_{a=1}^{k} p(l_a, s_1')$ but $u_1(s_1(D_1), s_2) \geq u_1(s_1'(D_1), s_2)$ for all valuations v_1. Consider a valuation v_1 where $v_1(l_k) = V > \epsilon > 0$, and with $V + \epsilon > v_i(l_a) > V$ for all $a < k$ and $\epsilon > v_i(l_a)$ for all $a > k$. For a contradiction, we want that

$$\sum_{a=1}^{k-1} p(l_a, s_1')v_1(l_a) + p(l_k, s_1')V + \sum_{a=k+1}^{d} p(l_k, s_1')v_1(l_a) >$$
$$\sum_{a=1}^{k-1} p(l_a, s_1)v_1(l_a) + p(l_k, s_1)V + \sum_{a=k+1}^{d} p(l_k, s_1')v_1(l_a)$$

For this, it sufficient to establish that

$$Vp(l_{\leq k}, s_1') > (V + \epsilon)p(l_{\leq k}, s_1) + \epsilon p(l_{>k}, s_1)$$

This is satisfied for any

$$V > \frac{\epsilon(p(l_{\leq k}, s_1) + p(l_{>k}, s_1))}{p(l_{\leq k}, s_1') - p(l_{\leq k}, s_1)},$$

from which we establish the contradiction.

■

Lemmas 2 and 3 can easily be extended to show that strict stochastic dominance implies strictly greater utility and vice versa, for all valuations consistent with match-early preferences.

Definition 7. *Second-stage strategy profile $s^* = (s_1^*, s_2^*)$ is a strict ordinal Bayesian-Nash equilibrium of the second-stage of the ESP game conditioned on effort levels e_1 and e_2 if for both players $i \in \{1, 2\}$, for every D_i, and every u_i consistent with match-early preferences,*

$$u_i(s_i^*(D_i), s_{-i}^*) > u_i(s_i'(D_i), s_{-i}^*), \tag{3}$$

where the probability adopted in interim utility u_i for the distribution on the dictionary of player $-i$ is induced by the effort of that player in the first stage.

Lemmas 2 and 3, taken together, establish that strict stochastic dominance of the distribution on outcomes induced by a strategy profile, in relation to the distribution on outcomes induced by a unilateral deviation to any other strategy, is necessary and sufficient for an ordinal BNE of the second-stage ESP game. The following is a simple technical lemma needed throughout for the equilibrium analysis.

Lemma 4. *If dictionary D and dictionary D' only differ by one element, x_i and x_i' respectively, with $f_e(x_i) < f_e(x_i')$, then dictionary D' is sampled with strictly greater probability than dictionary D under the same effort level e.*

Proof. $\Pr(D' \text{ is sampled}) = \sum_{\sigma(D')=x_1',x_2',..,x_d'} \Pr(x_1') \Pr(x_2'|x_1') ... \Pr(x_d'|x_1', x_2', ..., x_{d-1}')$, where we sum over all possible shufflings of the dictionary D', that is, all possible orderings in which the dictionary D' can be chosen. Likewise, $\Pr(D \text{ is sampled}) = \sum_{\sigma(D)=x_1,x_2,..,x_d} \Pr(x_1) \Pr(x_2|x_1) ... \Pr(x_d|x_1, x_2, ..., x_{d-1})$. Each shuffle $\sigma(D') = x_1', x_2', ..., x_d'$ has a corresponding shuffle $\sigma(D) = x_1, x_2, ..., x_d$, where $\sigma(D')$ and $\sigma(D)$ differ in the i^{th} coordinate only. Since $\sigma(D')$ and $\sigma(D)$ differ only in the i^{th} coordinate, $\Pr(x_1') \Pr(x_2'|x_1') ... \Pr(x_{i-1}'|x_1', x_2', ..., x_{i-2}') = \Pr(x_1) \Pr(x_2|x_1) ... \Pr(x_{i-1}|x_1, x_2, ..., x_{i-2})$. Since $f(x_i') > f(x_i)$, $\Pr(x_i'|x_1', x_2', ..., x_{i-1}') > \Pr(x_i|x_1, x_2, ..., x_{i-1})$. Now consider $\Pr(x_j'|x_1', x_2', ..., x_{j-1}')$ and $\Pr(x_j|x_1, x_2, ..., x_{j-1})$ for all $i < j \leq d$. We know that $f(x_j') = f(x_j)$ and $1 - f(x_1') - f(x_2') - ... - f(x_{j-1}') < 1 - f(x_1) - f(x_2) - ... - f(x_{j-1})$, so $\Pr(x_j'|x_1', x_2', ..., x_{j-1}') > \Pr(x_j|x_1, x_2, ..., x_{j-1})$ for all $i < j \leq d$. Since for each $\sigma(D')$, $\sigma(D)$ pair, we have established $\Pr(\sigma(D') \text{ is sampled in that order}) > \Pr(\sigma(D) \text{ is sampled in that order})$, we know that $\Pr(D' \text{ is sampled}) > \Pr(D \text{ is sampled})$. ■

In what follows, we show that "playing decreasing frequency" is a strict ordinal Bayesian-Nash equilibrium of the second-stage ESP game. Moreover, we show that this equilibrium is the only ordinal Bayesian-Nash equilibrium that holds for every distribution over universe U.

Below we describe a generalized strategy for player 1 in terms of player 2's second-level strategy s_2. We can compute a candidate best response for player 1 given her sampled dictionary D_1, the distribution over U (which induces a distribution over the opponent's dictionary D_2), her opponent's effort level e_2, and the second-level strategy s_2 of player 2. Note that this algorithm only gives a completely specified strategy if we run the algorithm for all possible dictionaries $D_1 \in \mathcal{D}_1$. When this algorithm is run for particular value of D_1, it outputs an *ordering* on the words in D_1. The candidate best response algorithm is given below.

Algorithm 1 implicitly takes into account the effort level of player 2. If player 2 is playing a lower effort level than player 1, player 1 will play those words in $D_1 \cap U_{e_2}$ followed by any words in D_1 that are not in U_{e_2} (these are the higher effort words that player 2 did not sample). Likewise, if player 2 is playing a higher effort level than player 1, this algorithm still gives the best response for player 1. Since the higher effort words that player 2 may have are not in her sampled dictionary, she cannot play them.

We say that the output of Algorithm 1 with respect to dictionary D is *consistent* with s_2 if for all pairs of words $w_i', w_j' \in D$, Algorithm 1 specifies playing w_i' before w_j' if and only if $w_i' \succ w_j'$ in s_2. If the output of Algorithm 1 with respect to dictionary D is not *consistent* with s_2, then we say it is *inconsistent* with s_2.

Algorithm 1 Candidate Best Response for Player 1

Input: sampled D_1, $\sigma_2 = (e_2, s_2)$
Maintain ordered list $\mathcal{O}_{BR} = \emptyset$
for $i = 1$ to d **do**
  Add element

$$E_{add} = \underset{w_j \in D_1 - \mathcal{O}_{BR}}{\arg\max} \sum_{D_2 \in \mathcal{D}_{e_2}} \Pr(D_2) \cdot I(w_j \text{ is in the top } i \text{ of } s_2(D_2))$$

  to the end of the ordered list $\mathcal{O}_{BR}$
end for
Output: $\mathcal{O}_{BR}$

It is important to note that this algorithm does not always output an ordering that stochastically dominates all other orderings, for a fixed D_1. But, when it does not output an ordering that stochastically dominates all other orderings, for a particular D_1, we show that no such ordering exists. In other words, this algorithm will output an ordering on D_1 that stochastically dominates all other orderings on D_1, when such an ordering exists.

The following definition is useful to us in characterizing the output of Algorithm 1. Note that the set $\{w'_1, ..., w'_n\}$ is ordered according to s_2 with and the set $\{w_1, w_2, ..., w_n\}$. We also use the notation that $w'_i \in l_k(s_2(D_2))$ means that word w'_i is the k^{th} highest priority word in dictionary D_2, when s_2 acts on D_2. Similarly, in the following definition $w'_i \in l_{\leq k}(s_2(D_2))$ means that word w'_i is among the k highest priority words of dictionary D_2.

Definition 8. *We say that second-stage strategy s_2 satisfies the preservation condition for a particular distribution, if for a fixed effort level of player 2 and for every pair of w'_i and w'_j such that $i < j$, we have that $\Pr(w'_i \in l_{\leq k}(s_2(D_2))) > \Pr(w'_j \in l_{\leq k}(s_2(D_2)))$ for all $1 \leq k \leq i$.*

Definition 9. *We say that s_2 satisfies the strong condition for a particular distribution, if for a fixed effort level and for every pair of w'_i and w'_j such that $i < j - 1$, we have that $\Pr(w'_i \in D_2) > \Pr(w'_j \in l_{\leq k}(s_2(D_2)))$ for all $i < k < j$.*

Lemma 5. *If Algorithm 1 outputs an ordering that does not stochastically dominate all other orderings, with respect to D_1 and for fixed opponent strategy σ_2, then no such ordering exists.*

Proof. Assume otherwise, that is, assume that there exists an ordering s' on D_1 that stochastically dominates all other possible strategies. This ordering $s'(D_1)$ is not output by Algorithm 1 by the statement of the theorem. Call the ordering output by Algorithm 1 $s(D_1)$. Consider the first coordinate in which $p(s(D_1), s_2)$ and $p(s'(D_1), s_2)$ differ, call this coordinate l_i. It must be the case that $p(l_i, s(D_1), s_2) \geq p(l_i, s'(D_1), s_2)$ since Algorithm 1 will output the word (of the remaining words) that will be the most likely to appear in the top i words of player 2 and this word is the first word in which the two strategies differ. Therefore, $\sum_{a=1}^{i} p(l_a, s(D_1), s_2) \geq \sum_{a=1}^{i} p(l_a, s'(D_1), s_2)$. ■

Recall from Lemma 3, that stochastic dominance is a necessary condition in order to have utility maximization for all valuations consistent with match early preferences.

The following lemma gives sufficient conditions for the strategy of player 1 to always output an ordering consistent with s_2 and for this strategy to stochastically dominate all other strategies.

Lemma 6. *If second-stage strategy s_2 satisfies the preservation condition for a particular distribution, then Algorithm 1 will always output an ordering consistent with s_2, for any sampled dictionary. Moreover, if s_2 satisfies the strong condition for this distribution, then the strategy of always playing an ordering consistent with s_2 will strictly stochastically dominate all other strategies, for any sampled dictionary.*

Proof. Since s_2 satisfies the preservation condition, we know that for every dictionary D and every pair of words $d_i, d_j \in D$ such that $d_i \succ d_j$ in the prior ordering, $\Pr(d_i \in l_{\leq k}(s_2(D_2))) > \Pr(d_j \in l_{\leq k}(s_2(D_2)))$ for all $1 \leq k \leq i$. Since $\Pr(d_i \in l_{\leq k}(s_2(D_2))) > \Pr(d_j \in l_{\leq k}(s_2(D_2)))$ for all $1 \leq k \leq i$, d_j cannot be output before d_i by Algorithm 1, for every dictionary D. Since this is true for all pairs of $d_i, d_j \in D$ such that $d_i \succ d_j$, Algorithm 1 must output an ordering consistent with s_2.

Now it remains to show that playing the strategy that is consistent with s_2 for all dictionaries stochastically dominates all other strategies. In order for a strategy to achieve stochastic dominance, we need the ordering $s(D)$ to stochastically dominate all other possible output orderings of D, for every possible dictionary D. In order for an ordering $s_1(D_1)$ to achieve stochastic dominance, we need the strategy to simultaneously satisfy the following conditions: the first word must maximize $p(l_1, s_1(D_1), s_2)$ (out of all possible output orderings of one word), the first two words played must maximize $p(l_1, s_1(D_1), s_2) + p(l_2, s_1(D_1), s_2)$ (out possible output orderings of two words), ..., the first i words played must maximize $\sum_{j=1}^{i} p(l_j, s_1(D_1), s_2)$ (out of all possible output orderings of i words), etc. For every dictionary D_1, where $D_1 = \{w''_1, w''_2, ..., w''_d\}$, where $w''_1 \succ w''_2 \succ ... \succ w''_d$ under s_2, we are given that $\Pr(d'_i \in D) > \Pr(d'_j \in D(k))$ for all i, j, k, such that $i < k < j$. Thus we know that w''_1 is the word that strictly maximizes $p(l_1, s_1(D_1), s_2)$, w''_1, w''_2 are the words that strictly maximize $p(l_1, s_1(D_1), s_2) + p(l_2, s_1(D_1), s_2)$, ..., w''_1, w''_2, ..., w''_i are the words that strictly maximize $\sum_{j=1}^{i} p(l_j, s_1(D_1), s_2)$, etc. ∎

Lemma 7. *The strategy profile $(s_1^{\downarrow}, s_1^{\downarrow})$ is a strict ordinal Bayesian-Nash equilibrium of the second-stage ESP game, for every choice of effort levels e_1 and e_2, for every distribution over U.*

Proof. For each pair of words w'_i, w'_j such that $w'_i \succ w'_j$ in $s_2^{\downarrow}$, we know that $f(w'_i) \geq f(w'_j)$ for all distributions over U. Consider the set of dictionaries that satisfy the property that $w'_i \in l_{\leq k}(s_2(D_2)) \cap w'_j \notin l_{\leq k}(s_2(D_2))$ and the set of dictionaries that satisfy the property that $w'_j \in l_{\leq k}(s_2(D_2)) \cap w'_i \notin l_{\leq k}(s_2(D_2))$, for any value of $1 \leq k \leq i$. Call the former set A and the latter set B. Notice that set B is the same as the set of dictionaries that satisfy the property that $w'_j \in l_{\leq k}(s_2(D_2)) \cap w'_i \notin D_2$. There exists a mapping $t : B \to A$, which takes an element $b \in B$ to an element $a \in B$, by removing w'_j and replacing it with w'_i. The mapping t takes each element $b \in B$ to a unique element in A (though not every element in A is hit by this mapping). Moreover, the mapping takes an element $b \in B$ to an element $a \in A$, where $\Pr(b) < \Pr(a)$, due to Lemma 4. Therefore $Pr(B) \leq Pr(A)$. Hence the *preservation condition* is satisfied when $s_2 = s_2^{\downarrow}$, regardless of distribution. Finally, when $s_2 = s_2^{\downarrow}$, we have that $\Pr(w'_i \in D) > \Pr(w'_j \in D)$, for all j, i such that $j > i$. Likewise, $\Pr(w'_j \in D) > \Pr(w'_j \in l_{\leq k}(s_2(D_2))$, for all j, k where $j > k$. This gives us: $\Pr(w'_i \in D) > Pr(w'_j \in D) > \Pr(w'_j \in l_{\leq k}(s_2(D_2)))$ for all i, j, k where $i < k < j$ and for all distributions or in other words, the *strong condition* is satisfied for all distributions. Therefore Lemma 6, along with Lemma 2, gives us the desired result. ∎

Lemma 8. *The strategy profile (s_2, s_2) is a strict ordinal Bayesian-Nash equilibrium of the second-stage ESP game for every s_2, for $e_1 = e_2$, for the uniform distribution over U.*

Proof. Since we have the uniform distribution over U, for all pairs of words w'_i, w'_j such that $w'_i \succ w'_j$ in s_2, $\Pr(w'_i \in D_2) = \Pr(w'_j \in D_2)$. This gives us that $\Pr(w'_i \in D_2 \cap w'_j \notin D_2) = \Pr(w'_j \in D_2 \cap w'_i \notin D_2)$. Likewise, under the uniform distribution, we have that $\Pr(w'_i \in l_{\leq k}(s_2(D_2)) \cap w'_j \notin l_{\leq k}(s_2(D_2))) > \Pr(w'_j \in l_{\leq k}(s_2(D_2)) \cap w'_i \notin l_{\leq k}(s_2(D_2)))$ for all $1 \leq k \leq i$. Thus s_2 satisfies the *preservation condition* under the uniform distribution, for every s_2. From Lemma 6 we get that s_2 is a best response to s_2. Finally, under the uniform distribution, for every s_2, we have that $\Pr(w'_i \in D) = \Pr(w'_j \in D)$, for all j, i such that $j > i$. Likewise, $\Pr(w'_j \in D) > \Pr(w'_j \in l_{\leq k}(s_2(D_2)))$, for all j, k where $j > k$. This gives us: $\Pr(w'_i \in D) = Pr(w'_j \in D) > \Pr(w'_j \in l_{\leq k}(s_2(D_2))$ for all i, j, k where $i < k < j$ or in other words, the *strong condition* is satisfied. Therefore Lemma 6, along with Lemma 2, gives us the desired result. ■

In other words, if we had a strategy profile (s'_1, s_2), for some $s'_1 \neq s_2$, then Lemma 8 tells us that this strategy profile cannot be an ordinal Bayesian-Nash equilibrium under the uniform distribution, because for some distribution over U and some sampled dictionary, s_2 generates a distribution on outcomes that strictly stochastically dominates $s'_1 \neq s_2$. Therefore, Lemma 8 tells us that if a strategy profile is an ordinal Bayesian-Nash equilibrium for all distributions over U, then it must be that case that this strategy profile is symmetric.

It should be noted that the statement of the above lemma can be slightly generalized to take care of the case where players play different effort levels, but still under the uniform distribution over the words in U. If player 1 is playing a lower effort level than player 2, (s'_2, s_2) is a strict Bayesian-Nash equilibrium for every s_2, where s_2 is a total ordering on the set U_{e_2} and s'_2 is s_2 with all the words in the set $U_{e_2} - U_{e_1}$ removed. Likewise, if player 2 is playing a lower effort level than player 1, (s'_2, s_2) is a strict Bayesian-Nash equilibrium for every s_2, where s_2 is a total ordering on the set U_{e_2} and s'_2 is s_2 with all the words in the set $U_{e_1} - U_{e_2}$ concatenated to the end of $\mathcal{O}_{prior}$, e.g. all words in $U_{e_1} - U_{e_2}$ are lower priority than all words in U_{e_2} under $\mathcal{O}'_{prior}$.

Lemma 9. *For each s_2 (except for $s_2^{\downarrow}$), there exists a distribution and a dictionary for which Algorithm 1 will not always output an ordering consistent with s_2.*

Proof. Let s_2 be: $w'_1 \succ w'_2 \succ ... \succ w'_{|U_e|}$. Since $s_2 \neq s_2^{\downarrow}$, there exists a pair of adjacent words that are "out-of-order". That is for some adjacent pair of words, w'_i, w'_{i+1} where $w'_i \succ w'_{i+1}$, we have $f(w'_i) < f(w'_{i+1})$. Consider the smallest value of i such that the preservation condition is violated and consider any dictionary that contains the highest $i+1$ priority words, according to s_2. Since the preservation condition is satisfied by s_2, when you restrict attention to the first i words, we know from Lemma 6, that Algorithm 1 will output these $i-1$ words in the first $i-1$ steps of the algorithm. It suffices to show that there exists a distribution for which $\Pr(w'_i \in l_{\leq i}(s_2(D_2))) < \Pr(w'_{i+1} \in l_{\leq i}(s_2(D_2)))$, because if this is the case, then w'_i cannot be output at step i of the algorithm. This is the same as satisfying $\Pr(w'_i \in l_{\leq i}(s_2(D_2)) \cap w'_{i+1} \notin l_{\leq i}(s_2(D_2))) < \Pr(w'_{i+1} \in l_{\leq i}(s_2(D_2)) \cap w'_i \notin l_{\leq i}(s_2(D_2)))$, or in other words, $\Pr(w'_i \in D \cap w'_{i+1} \notin D) + \Pr(w'_i \in l_i(D) \cap w'_{i+1} \in l_{i+1}(D)) < \Pr(w'_{i+1} \in D \cap w'_i \notin D)$. Thus it suffices to show that $\Pr(w'_i \in D) < \Pr(w'_{i+1} \in D \cap w'_i \notin D)$. This is the same as $\Pr(w'_i \in D) < \Pr(w'_i \notin D) \cdot \Pr(w'_{i+1} \in D | w'_i \notin D)$, or in other words $\Pr(w'_{i+1} \in D | w'_i \notin D) > \frac{\Pr(w'_i \in D)}{\Pr(w'_i \notin D)}$. We can make $f(w'_i)$ arbitrarily small, so that $\Pr(w'_i \in D)$ is also arbitrarily small, that is, say ϵ and we can easily choose a value of $f(w'_{i+1})$ such that $\Pr(w'_{i+1} \in D) > \frac{\epsilon}{1-\epsilon}$, which gives us the desired property. ■

Theorem 1. *Second-stage strategy profile $(s_1^{\downarrow}, s_2^{\downarrow})$ is a strict ordinal Bayesian-Nash equilibrium for the second-stage ESP game for every distribution over U. Moreover, the $(s_1^{\downarrow}, s_2^{\downarrow})$ strategy profile is the only strategy profile in which at least one player plays a consistent strategy that can be an ordinal Bayesian-Nash equilibrium for every distribution over U.*

Proof. Lemma 7 tells us that Algorithm 1 will always output a strategy consistent with $s_2^{\downarrow}$ if player 2 is playing $s_2^{\downarrow}$, regardless of the dictionary D_1 and regardless of the distribution over U. Furthermore, this strategy stochastically dominates all other strategies, for every distribution over U. Lemma 8 tells us that there exists a distribution, namely the uniform distribution, for which Algorithm 1 will output an ordering consistent with s_2, regardless of the dictionary D_1, for all s_2. Moreover, this strategy will stochastically dominate all others, for the uniform distribution. Lemma 9 tells us that there exists a distribution $F(U)$ and dictionary D_1 for which Algorithm 1 will output an ordering inconsistent with s_2, for all $s_2 \neq s_2^{\downarrow}$. Either this strategy stochastically dominates all others, for this distribution $F(U)$, or it does not. In the former case, we have exhibited two distributions that have two different strategies that stochastically dominate all others. In the latter case, we know that there is no strategy that stochastically dominates all others for the distribution $F(U)$ from Lemma 5. Therefore, there is no one strategy for player 1 that stochastically dominates all others when player 2 is playing $s_2 \neq s_2^{\downarrow}$ for all distributions over U and every valuation function that satisfies match-early preferences. ■

We have shown that the strategy profile $(s_1^{\downarrow}, s_2^{\downarrow})$ is an ordinal Bayesian-Nash equilibrium of the second-stage of the ESP game for every distribution over U. Moreover, we have shown that this is the only strategy profile, where at least one player is playing a consistent strategy, that is an ordinal Bayesian-Nash equilibrium for every distribution over U. Therefore, we focus on the $(s_1^{\downarrow}, s_2^{\downarrow})$ equilibrium profile when analyzing the equilibrium of the complete game.

In the results that follow, we show that playing L at the top-level together with playing words in order of decreasing frequency is an ordinal Bayesian-Nash equilibrium. In order to show this, we use the following definition of stochastic dominance for the top level of the game which fixes the equilibrium strategy for the bottom-level. The definition uses the following notation for a k-truncation of dictionary D: $D(k)$ is the set of k highest frequency words in D.

Definition 10. *Fixing player 2's strategy (e_2, s_2), we say that strategy (e_1, s_1) for player 1 stochastically dominates strategy (e_1', s_1) for player 1 if and only if:*

$$\sum_{D_{1,e_1}} \Pr(D_{1,e_1}|e_1) \sum_{D_{2,e_2}} \Pr(D_{2,e_2}|e_2) \cdot I(g_l(s_1(D_{1,e_1}(k)), s_2(D_{2,e_2}(k))) = l_1, ..., l_k) \geq$$
$$\sum_{D_{1,e_1'}} \Pr(D_{1,e_1'}|e_1') \sum_{D_{2,e_2}} \Pr(D_{2,e_2}|e_2) \cdot I(g_l(s_1(D_{1,e_1'}(k)), s_2(D_{2,e_2}(k))) = l_1, ..., l_k) \; \forall k$$

where $g_l(s_1(D_{1,e_1}(k)), s_2(D_{2,e_2}(k)))$ gives the outcome when second-stage strategies s_1 and s_2 act on truncated dictionaries $D_{1,e_1}(k)$ and $D_{2,e_2}(k)$ and $I(\cdot)$ is the indicator function. We say the stochastic dominance is strict if there exists a k such that the above inequality is strict.

Since Lemma 7 establishes that $(s_1^{\downarrow}, s_2^{\downarrow})$ is a strict ordinal Bayesian-Nash equilibrium of the second-stage ESP game, for all effort levels, we set $(s_1, s_2) = (s_1^{\downarrow}, s_2^{\downarrow})$ and we know that $I(g_l(s_1(D_{1,e_1}(k)), s_2(D_{2,e_2}(k))) = l_1, ..., l_k) = I(D_{1,e_1}(k) \cap D_{2,e_2}(k) \neq \emptyset)$ since the expression on the left hand side is simply the probability that a match occurs in the first k locations given that $(s_1, s_2) = (s_1^{\downarrow}, s_2^{\downarrow})$, which is exactly the probability that player 1's "top k" words overlap with player 2's "top k" words. The following two lemmas are the exact equivalents of Lemmas 2 and 3, this time for the full game. Likewise, these lemmas can easily be extended to show strict stochastic dominance implies strictly greater utility and vice versa, for all valuations consistent with match-early preferences.

Lemma 10. *If strategy $\sigma_1 = (e_1, s_1)$ stochastically dominates strategy $\sigma_1' = (e_1', s_1)$, for fixed opponent strategy $\sigma_2 = (e_2, s_2)$, then $u_1(\sigma_1, \sigma_2) \geq u_1(\sigma_1', \sigma_2)$, for all valuations consistent with match-early preferences.*

Lemma 11. *If $u_1(\sigma_1, \sigma_2) \geq u_1(\sigma_1', \sigma_2)$ (where $\sigma_1 = (e_1, s_1)$, $\sigma_1' = (e_1', s_1)$, and $\sigma_2 = (e_2, s_2)$) for all valuations that are consistent with match-early preferences, then strategy σ_1 must stochastically dominate strategy σ_1' for fixed opponent's strategy σ_2.*

In order to establish stochastic dominance, we construct a randomized mapping for each dictionary that can be sampled when playing M to a number of dictionaries that can be sampled when playing L. Each dictionary in $\mathcal{D}_M$ is mapped to a dictionary in $\mathcal{D}_L$ that is at least as likely to match against the opponent's dictionary, averaged over the distribution of all possible dictionaries for the opponent. This is shown in Lemma 13. In order to complete the proof, it is necessary to show that under the randomized mapping, no element in $\mathcal{D}_L$ is mapped to with greater probability under the randomized mapping than under the original distribution over $\mathcal{D}_L$. This fact is shown in Lemma 14.

We say that dictionary D' with elements $\{w'_1, w'_2, ..., w'_n\}$ (in order of decreasing frequency) *dominates* dictionary D with elements $\{w_1, w_2, ..., w_n\}$ (in order of decreasing frequency) if $f(w'_i) \geq f(w_i)$ for all i. We say that the dominance is *strict* if $D' \neq D$. The following lemma is need to prove Lemma 13.

Lemma 12. *For every pair of dictionaries D' and D such that dictionary D' dominates dictionary D, every effort level of player 2 and when both players play decreasing frequency in the second stage, we have that:*

$$\sum_{D_2} \Pr(D_2) \cdot I(D'(k) \cap D_2(k) \neq \emptyset) \geq$$
$$\sum_{D_2} \Pr(D_2) \cdot I(D(k) \cap D_2(k) \neq \emptyset) \; \forall k \tag{4}$$

In addition, when D' strictly dominates D, the inequality is strict for all $k \geq k'$, where k' is the first coordinate where D' and D differ.

Proof. It suffices to show equation 4 for dictionaries $D' = \{w'_1, w'_2, ..., w'_n\}$ (in sorted order) and $D = \{w_1, w_2, ..., w_n\}$ (in sorted order) that are identical except in the i^{th} coordinate where $f(w'_i) > f(w_i)$. If $k < i$, equation 4 holds with equality. Consider $k \geq i$ and let $A = \{w_1, w_2, ..., w_k\} - \{w_i\}$. $\sum_{D_2} \Pr(D_2) I(D'(k) \cap D_2(k) \neq \emptyset) = \Pr(D_2(k) \cap A \neq \emptyset) + \Pr((D_2(k) \cap A = \emptyset) \cup (w'_i \in D_2(k)))$.
Likewise, $\sum_{D_2} \Pr(D_2) I(D(k) \cap D_2(k) \neq \emptyset) = \Pr(D_2(k) \cap A \neq \emptyset) + \Pr((D_2(k) \cap A = \emptyset) \cup (w_i \in D_2(k)))$. We say that D satisfies property P if $D(k) \cap A = \emptyset$ and $w_i \in D(k)$. Likewise, we say that D satisfies property P' if $D(k) \cap A = \emptyset$ and $w'_i \in D(k)$.
We claim that $\Pr(D_2 \text{ satisfies } P') > \Pr(D_2 \text{ satisfies } P)$ by the following transformation $t : D_i \rightarrow D'_i$: Consider any dictionary D_i that satisfies property P. If D_i also has w'_i, the transformation t leaves the dictionary untouched and note the D'_i is a dictionary that satisfies property P'. Since $D_i = D'_i$, the probability that D_i is sampled is the same as the probability that D'_i is sampled. If D_i does not have w'_i, the transformation replaces w_i with w'_i to yield dictionary D'_i. Since $w_i \in D_i(k)$, $w'_i \in D'_i(k)$. Likewise, since $D_i(k) \cap A = \emptyset$, $D'_i(k) \cap A = \emptyset$. By Lemma 4, the probability of sampling dictionary D'_i is strictly greater than the probability of sampling dictionary D_i. Thus, we have established that: $\Pr(D_2 \text{ satisfies } P') > \Pr(D_2 \text{ satisfies } P)$. ■

For the following lemmas we use the randomized mapping h: Consider a dictionary $D \in \mathcal{D}_M$, $D = A \cup B$, where A is the set of "low words" and B is the set of "medium words" (in other words, $A = D \cap U_L$ and $B = D \cap (U_M - U_L)$). Under our randomized mapping, D is mapped to all dictionaries in $D_L \in \mathcal{D}_L$ such that $A \subset D_L$. In other words, D is mapped to dictionary $D_L \in \mathcal{D}_L$ with non-zero probability if and only if $A \subset D_L$. If $A \subset D_L$, then D is mapped to D_L with the same probability that you could would get D_L if you continued to sample individual words from U_M (without replacement) until you got d "low words". Note that if D contains only medium words, D is mapped to all dictionaries in $\mathcal{D}_L$ with non-zero probability. Likewise, if D contains only low words, D is mapped to only one dictionary in $\mathcal{D}_L$.

Lemma 13. *For every $D_{1,M}$, where $D_{1,M}$ is a dictionary sampled with respect to the M effort level, and for every h that satisfies the property that $D_{1,M}$ is mapped to a dictionary in $\mathcal{D}_L$ that contains the set $D_{1,M} \cap U_L$*

and every effort level of player 2 and when both players play decreasing frequency in the second stage, we have that:

$$\sum_{D_2} \Pr(D_2) \cdot I(h(D_{1,M})(k) \cap D_2(k) \neq \emptyset) \geq \sum_{D_2} \Pr(D_2) \cdot I(D_{1,M}(k) \cap D_2(k) \neq \emptyset) \ \forall k \text{ and } D_{1,M} \tag{5}$$

In addition, the inequality is strict for all $k \geq k'$ when $h(D_{1,M}) \neq D_{1,M}$ and k' is the first coordinate where $h(D_{1,M})$ and $D_{1,M}$ differ.

Proof. Due to lemma 12, it suffices to show that dictionary $h(D_{1,M}) = \{w'_1, w'_2, ..., w'_d\}$ dominates dictionary $D_{1,M} = \{w_1, w_2, ..., w_d\}$. Assume otherwise, that is, assume there exists a coordinate i such that $f(w'_i) < f(w_i)$. Let j be the minimum such coordinate. $w_j \in U_L$ since $h(D_{1,M})$ contains words only in U_L. This means that $w_j \in h(D_{1,M})$. Since the dictionaries are in sorted order, this means that $w_j = w'_k$ for some $k < j$, however, this means that $h(D_{1,M})$ does not contain all of $D_{1,M} \cap U_L$, a contradiction. When $h(D_{1,M}) \neq D_{1,M}$, the dominance is strict. ■

Lemma 14 states that the distribution obtained from sampling U_L directly is the same as the distribution obtained from sampling a medium dictionary, followed by the randomized mapping (i.e. sampling U_M until you get d low words). The proof is easy and omitted.

Lemma 14. $\Pr(D_{1,L}|L) = \sum_{D_{1,M}} \Pr(D_{1,M}|M) \cdot \Pr(h(D_{1,M}) = D_{1,L})$

Lemma 15 uses Lemmas 13 and 14 to show that playing L stochastically dominates playing M, assuming players play decreasing frequency in the second stage. An identical argument can be used to show that playing L stochastically dominates playing H, assuming players play decreasing frequency in the second stage. It is also important to note that this argument is independent of the number of effort levels so the equilibrium analysis holds as we vary the number of effort levels, as long as there are at least two.

Lemma 15. *For every effort level of player 2 and when players play decreasing frequency in the second stage:*

$$\sum_{D_{1,L}} \Pr(D_{1,L}|L) \sum_{D_2} \Pr(D_2) \cdot I(D_{1,L}(k) \cap D_2(k) \neq \emptyset) > \sum_{D_{1,M}} \Pr(D_{1,M}|M) \sum_{D_2} \Pr(D_2) \cdot I(D_{1,M}(k) \cap D_2(k) \neq \emptyset) \ \forall k \tag{6}$$

Proof. From Lemma 14, we know that:

$$\sum_{D_{1,L}} \Pr(D_{1,L}) \sum_{D_2} \Pr(D_2) \cdot I(D_{1,L}(k) \cap D_2(k) \neq \emptyset) = \sum_{D_{1,L}} [\sum_{D_{1,M}} \Pr(D_{1,M}|M) \cdot \Pr(h(D_{1,M}) = D_{1,L})] \cdot \sum_{D_2} \Pr(D_2) \cdot I(D_{1,L}(k) \cap D_2(k) \neq \emptyset) \ \forall k$$

This is equivalent to writing:

$$
\begin{aligned}
&\sum_{D_{1,L}} \Pr(D_{1,L}) \sum_{D_2} \Pr(D_2) \cdot I(D_{1,L}(k) \cap D_2(k) \neq \emptyset) \\
&\quad > \sum_{D_{1,L}} [\sum_{D_{1,M}} \Pr(D_{1,M}|M) \cdot \Pr(h(D_{1,M}) = D_{1,L}) \cdot \sum_{D_2} \Pr(D_2) \cdot I(h(D_{1,M})(k) \cap D_2(k) \neq \emptyset)] \\
&\quad \geq \sum_{D_{1,L}} [\sum_{D_{1,M}} \Pr(D_{1,M}|M) \cdot \Pr(h(D_{1,M}) = D_{1,L}) \cdot \sum_{D_2} \Pr(D_2) \cdot I(D_{1,M}(k) \cap D_2(k) \neq \emptyset)] \\
&\quad = \sum_{D_{1,M}} \Pr(D_{1,M}|M) \cdot \sum_{D_2} \Pr(D_2) \cdot I(D_{1,M}(k) \cap D_2(k) \neq \emptyset) \; \forall k
\end{aligned}
$$

where the last inequality follows from Lemma 13. We know that the first inequality is strict for all k since there exists a $D_{1,M}$ such that $h(D_{1,M}) \neq D_{1,M}$ and $h(D_{1,M})$ and $D_{1,M}$ differ in the first coordinate, for all possible values of $h(D_{1,M})$ under the randomized mapping. ■

Theorem 7 together with Lemmas 15 and 10 give us the following result.

Theorem 2. *$((L, s_1^{\downarrow}), (L, s_2^{\downarrow}))$ is a strict ordinal Bayesian-Nash equilibrium for the complete ESP game for every distribution over U. Moreover, $(L, s_1^{\downarrow})$ is a strict best-response to both $(M, s_2^{\downarrow})$ and $(H, s_2^{\downarrow})$ for every valuation function that satisfies match-early preferences and for every distribution over U.*

5 The Effect of Rare-Words First Preferences

In this section, we consider the effect of modified preferences. We introduce a new model called *rare-words first* preferences. Our motivation is to consider the effect of alternate payoff structures and in particular to understand whether there is a *high effort* equilibrium available if players care first about matching on rare words and then about matching early rather than late. We show that *decreasing frequency* is no longer stable with respect to stochastic dominance in the second-stage of the ESP game, whatever the distribution on words in the universe. This in direct contrast to the match-early preferences, where decreasing frequency is a strict Bayesian-Nash equilibrium of the second stage conditioned on any effort levels and for any underlying distribution and any valuations consistent with match-early preferences. We also impose more structure on the valuation models and find that playing words in order of *increasing frequency* can be a Bayesian-Nash equilibrium in both preference models, however the rare-word-first model supports this for a broader range of distributions than the match-early preferences. First we formally define *rare-words first* preferences.

Definition 11. *Under rare-words first preferences, players prefer to match on rare words, with location as a secondary consideration. Any valuation function $v(o)$ that satisfies rare-words first preferences satisfies the following total ordering on outcomes: $(w_n, l_1) \succ (w_n, l_2) \succ ... \succ (w_1, l_{d-1}) \succ (w_1, l_d) \succ (\phi, \phi)$.*

Much like the previous section, we have an analogous "probability of a first match" vector. For each word w_i, a second-stage strategy s_1 acting on dictionary D_1 (given that opponent plays second-stage strategy s_2) gives a probability vector, $\vec{p}(w_i, s_1(D_1), s_2)$ = $(p(w_i, l_1, s_1(D_1), s_2)$, $p(w_i, l_2, s_1(D_1), s_2)$, ..., $p(w_i, l_d, s_1(D_1), s_2))$, of matching with the opponent on that given word (summed over all possible dictionaries of the opponent). Entry $p(w_i, l_j, s_1(D_1), s_2)$ in the probability vector is the probability of matching in location j on word i. We concatenate the probability vectors for each word, to get a single vector, $\mathcal{V}$, where the entries of $\mathcal{V}$ are ordered according to the valuation function $v(o)$. In other words, $\vec{\mathcal{V}}(s_1(D_1), s_2)$ = $(p(w_n, l_1, s_1(D_1), s_2)$, $p(w_n, l_2, s_1(D_1), s_2)$, ..., $p(w_1, l_{d-1}, s_1(D_1), s_2)$,

$p(w_1, l_d, s_1(D_1), s_2)$, $p(\phi, \phi, s_1(D_1), s_2))$. Our definition of stochastic dominance for this section is virtually identical to that of the previous section along with the results that the stochastic dominance condition is both sufficient and necessary.

Definition 12. *For two vectors $\vec{V}$ and $\vec{V'}$ of length m, we say that vector $\vec{V}$ stochastically dominates vector $\vec{V'}$ if and only i f $\sum_{a=1}^{k} \vec{V}(a) \geq \sum_{a=1}^{k} \vec{V'}(a)$.*

Definition 13. *Fixing opponent's second-stage strategy s_2, we say that second-stage strategy s_1 stochastically dominates second-stage strategy s'_1 with respect to dictionary D_1 if vector $\vec{\mathcal{V}}(s_1(D_1), s_2)$ stochastically dominates $\vec{\mathcal{V}}(s'_1(D_1), s_2)$ according to Definition 12.*

Lemma 16. *If second-stage strategy s_1 stochastically dominates second-stage strategy s'_1 with respect to dictionary D_1, for fixed opponent second-stage strategy s_2, then $u_1(s_1(D_1), s_2) > u_1(s'_1(D_1), s_2)$ for all valuations consistent with rare-words first preferences.*

Lemma 17. *If $u_1(s_1(D_1), s_2) > u_1(s'_1(D_1), s_2)$ for all valuations that satisfy rare-words first preferences, then second-stage strategy s_1 stochastically dominates second-stage strategy s_2 with respect to dictionary D_1, for fixed opponent second stage strategy s_2.*

The following lemma is in stark contrast with the results in section 4, where we showed that $(\downarrow, \downarrow)$ is a strict Bayesian-Nash equilibrium in the second stage, for all distributions over U, all valuation functions that satisfy MEP, and any pair of effort levels. Lemma 18 shows that we cannot say $(\downarrow, \downarrow)$ is a Bayesian-Nash equilibrium for the second stage of the game for any distribution, without making more assumptions on the valuation function.

Lemma 18. *Consider any distribution over $U = \{w_1, w_2, ..., w_n\}$ and suppose that player 2 is playing her words in order of decreasing frequency. For any dictionary of player 1, no consistent strategy of player 1 can stochastically dominate all other consistent strategies under Definition 13.*

Proof. We present a simple example to illustrate the intuition behind this lemma, which can be easily generalized to get the statement of the lemma. Suppose we have $U = \{w_1, w_2, w_3, w_4\}$ with $d = 2$. Suppose player 1 has dictionary w_1, w_2. She can either play w_1 followed by w_2 (call this strategy s_1) or w_2 followed by w_1 (call this strategy s'_1). If player 1 is playing strategy s_1, she will match with player 2 on w_1 in location 1 when player 2 has dictionary w_1, w_2 or w_1, w_3 or w_1, w_4. She will match with player 2 on w_2 in location 2 when player has dictionary w_2, w_3 or w_2, w_4. This leads to a match vector of $\vec{v} = (0, x, y + z, 0)$ where x is the probability that player 2 has either w_2, w_3 or w_2, w_4 as her dictionary, y is the probability that player 2 has either w_1, w_2 as her dictionary, and z is the probability that player 2 has w_1, w_3 or w_1, w_4 as her dictionary. If player 1 is playing strategy s'_1, she will match with player 2 on w_2 in location 1 when player 2 has dictionary w_2, w_3 or w_2, w_4 and on w_2 in location 2 when player 2 has dictionary w_1, w_2 (since we break ties in favor of the lower frequency word). She will match with player 2 on w_1 in location 2 when player has dictionary w_1, w_3 or w_3, w_2. This leads to a match vector of $\vec{v}' = (x, y, 0, z)$. It is easy to see that neither $\vec{v}$ nor $\vec{v}'$ stochastically dominates the other for any $x, y, z > 0$. Since $x, y, z > 0$ for any distribution over U that assigns non-zero probability to each of the elements in U, neither s_1 nor s'_1 can stochastically dominate the other for any distribution over U. ■

Similarly, Lemma 19 shows that when a player is playing *increasing frequency*, we need to make more assumptions on the valuation function to discern the best-response in the space of consistent strategies.

Lemma 19. *Consider any distribution over $U = \{w_1, w_2, ..., w_n\}$ and suppose that player 2 is playing her words in order of increasing frequency. For any dictionary of player 1, no consistent strategy of player 1 can stochastically dominate all other consistent strategies under Definition 13.*

Proof. While we can show the general version of this lemma, we give the following simple example to illustrate the intuition behind the lemma. Suppose we have $U = \{w_1, w_2, w_3, w_4\}$ with $d = 2$. Suppose player 1 has dictionary w_3, w_4. She can either play w_3 followed by w_4 (call this strategy s_1) or w_4 followed by w_3 (call this strategy s_1'). If player 1 is playing strategy s_1, she will match with player 2 on w_3 in location 1 when player 2 has dictionary w_3, w_1 or w_3, w_2. She will match with player 2 on w_4 in location 2 when player has dictionary w_4, w_1 or w_4, w_2 or w_4, w_3 (we break "ties" with respect to the lower frequency word). This leads to a match vector of $v = (0, x, y, 0)$ where x is the probability that player 2 has either w_4, w_1 or w_4, w_2 or w_4, w_3 as her dictionary and y is the probability that player 2 has either w_3, w_1 or w_3, w_2 as her dictionary. If player 1 is playing strategy s_1', she will match with player 2 on w_4 in location 1 when player 2 has dictionary w_4, w_1 or w_4, w_2 or w_4, w_3. She will match with player 2 on w_3 in location 2 when player has dictionary w_3, w_1 or w_3, w_2. This leads to a match vector of $v' = (x, 0, 0, y)$. It is easy to see that neither v nor v' stochastically dominates the other for any $x, y > 0$. Since $x, y > 0$ for any distribution over U that assigns non-zero probability to each of the elements in U, neither s_1 nor s_1' can stochastically dominate the other for any distribution over U. ■

Since we have shown that assumptions on the valuation function are necessary, we introduce a natural constraint on the valuation function below.

Definition 14. *Additive Discount Property: The valuation function $v_i(o)$ over the total ordering of outcomes $o_1 \succ o_2 \succ \ldots \succ o_{nd-1} \succ o_{nd} \succ o_{nd+1}$ satisfies the additive discount property if and only if, for each pair of adjacent outcomes o_j and o_{j+1}, $v(o_j) - v(o_{j+1}) \geq \alpha$ for some constant $\alpha > 0$.*

Lemma 20. *Playing words in order of increasing frequency is a strict Bayesian-Nash equilibrium of the second level of the ESP game for any valuation that satisfies rare-words first preferences with additive discounting if* $\Pr(w_i \notin l_{\leq k}(D_2) \cap w_j \in l_{\leq k}(D_2)) + \Pr(w_i \in l_{\leq k}(D_2) \cap w_j \in l_{\leq k}(D_2)) \cdot (d-1) > \Pr(w_j \notin D_2 \cap w_i \in l_{\leq k}(D_2))$ *for all pairs of words $w_i, w_j \in U$ that satisfy $i < j$ (in other words $f(w_i) > f(w_j)$) and all $k \leq |U_{e_1}| - j + 1$, given that $e_1 \leq e_2$.*

Proof. Suppose that player 2 is playing his words in order of increasing frequency (call this strategy s_2) and player 1 has type D_1 and is playing strategy $s_1(D_1) \rightarrow w_1' \succ w_2' \succ \ldots \succ w_d'$ (which is the ordering specified by strategy s_1 acting upon D_1). We show that if there exists a pair of adjacent words $w_i' \succ w_{i+1}'$ in $s_1(D_1)$ such that $f(w_i') > f(w_{i+1}')$, the ordering $s_1'(D_1)$ that is identical to $s_1(D_1)$ except with $w_{i+1}' \succ w_i'$ yields strictly greater utility than $s_1(D_1)$. We show that $\sum_{D_2}(u_1(s_1'(D_1), s_2(D_2)) - u_1(s_1(D_1), s_2(D_2))) > 0$. For this summation, it suffices to restrict our attention to D_2 that contain either w_i' or w_{i+1}', because if D_2 does not contain either, swapping w_i' and w_{i+1}' does not change the outcome. (We say that in the ordering $s_1(D_1)$, word w_i' is in location i). If the match between $s_1(D_1)$ and $s_2(D_2)$ occurs before location i, swapping w_i' and w_{i+1}' does not change the outcome. Likewise, if the match between $s_1(D_1)$ and $s_2(D_2)$ occurs after location $i+1$, swapping w_i' and w_{i+1}' does not change the outcome. So we can restrict our attention to the dictionaries D_2 such that $s_2(D_2)$ matches with $s_1(D_1)$ in the i^{th} or the $i+1^{st}$ location and that contain w_i' or w_{i+1}'. Since the match occurs in the i^{th} or the $i+1^{st}$ location it either involves player 2's i^{th} or $i+1^{st}$ word and/or player 1's i^{th} or $i+1^{st}$ word (namely w_i' and w_{i+1}'). Consider the following cases:
(1) Suppose the match between $s_1(D_1)$ and $s_2(D_2)$ occurs on w_i' in location i (this means w_i' is in location $1, \ldots, i$ of D_2). If $w_{i+1}' \notin D_2$, swapping w_{i+1}' and w_i' leads to a less desirable outcome, namely, instead of matching on w_i' in location i, now the match occurs on w_i' in location $i+1$. If $w_{i+1}' \in D_2$, it is played before w_i' by player 1, so swapping w_{i+1}' and w_i' leads to a more desirable outcome, namely, instead of matching on w_i' in location i, now the match occurs on w_{i+1}' in location i.
(2) Suppose the match between $s_1(D_1)$ and $s_2(D_2)$ occurs on w_{i+1}' in location $i+1$ (this means w_i' is not in location $1, \ldots, i$ of D_2). If w_{i+1}' is in location $1, \ldots, i$ of D_2, then swapping w_{i+1}' and w_i' leads to a more desirable outcome, instead of matching on w_{i+1}' in location $i+1$, now the match occurs on w_{i+1}' in location

$i+1$. If w'_{i+1} is in location $i+1$ of D_2, swapping does not change the outcome.
(3) Suppose the match between $s_1(D_1)$ and $s_2(D_2)$ occurs on some $w \neq w'_i, w'_{i+1}$ in location i. This means w is in location i of D_2. If swapping w'_{i+1} and w'_i changes the outcome, the new match occurs on w'_{i+1}. Since there is a match on w'_{i+1} in location i and a match on w in location i after swapping, $f(w'_{i+1}) < f(w)$. This leads to a more favorable outcome, instead of matching on w in location i, the new match occurs on w_{i+1} in location i. This occurs when w_{i+1} is in location $1, ..., i-1$ of D_2.
(4) Suppose the match between $s_1(D_1)$ and $s_2(D_2)$ occurs on some $w \neq w'_i, w'_{i+1}$ in location $i+1$. We claim that swapping w'_{i+1} and w'_i does not change the outcome in this case. Assume otherwise, this means the match now occurs on w'_{i+1} in location i. Since the match did not occur on w'_{i+1} before the swap, it must be that $f(w'_{i+1}) > f(w)$. Since a match now occurs on w'_{i+1} in location i, it means that w'_{i+1} occurred in location $1, ..., i$ of D_2, which implies that $f(w'_{i+1}) < f(w)$, a contradiction.
Putting this together, we get: $\sum_{D_2}(u_1(s'_1(D_1), s_2(D_2)) - u_1(s_1(D_1), s_2(D_2)))$
$> \Pr(w'_i \notin l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2))(v(w'_{i+1}, l_i) - v(w'_{i+1}, l_{i+1})) - \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \notin D_2)(v(w'_i, l_i) - v(w'_i, l_{i+1})) + \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2))(v(w'_{i+1}, l_i) - v(w'_i, l_i))$
$> \Pr(w'_i \notin l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2)) \cdot \alpha - \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \notin D_2) \cdot \alpha + \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2)) \cdot (d-1) \cdot \alpha$, which is greater than 0 by the statement of the lemma. ■

Definition 15. *Multiplicative Discount Property: The valuation function $v_i(o)$ over the total ordering of outcomes $o_1 \succ o_2 \succ ... \succ o_{nd-1} \succ o_{nd} \succ o_{nd+1}$ satisfies the additive discount property if and only if, for each pair of adjacent outcomes o_j and o_{j+1}, $\frac{v(o_j)}{v(o_{j+1})} \geq \alpha$ for some constant $\alpha > 1$.*

Lemma 21. *Playing words in order of increasing frequency is a strict Bayesian-Nash equilibrium of the second level of the ESP game for any valuation that satisfies rare-words first preferences with multiplicative discounting if $\Pr(w_j \in l_{\leq k}(D_2)) \cdot (\alpha^d - \alpha^{d-1}) + \Pr(w_i \in l_{\leq k}(D_2) \cap w_j \in l_{\leq k}(D_2)) \cdot (\alpha^d - \alpha) > \Pr(w_j \notin D_2 \cap w_i \in l_{\leq k}(D_2)) \cdot (\alpha - 1)$ for all pairs of words $w_i, w_j \in U$ that satisfy $i < j$ (in other words $f(w_i) > f(w_j)$) and all $k \leq |U_{e_1}| - j + 1$, given that $e_1 \leq e_2$.*

Proof. The proof is virtually identical to that of Lemma 20, but with a different valuation model: $\sum_{D_2}(u_1(s'_1(D_1), s_2(D_2)) - u_1(s_1(D_1), s_2(D_2)))$
$> \Pr(w'_i \notin l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2))(v(w'_{i+1}, l_i) - v(w'_{i+1}, l_{i+1})) - \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \notin D_2)(v(w'_i, l_i) - v(w'_i, l_{i+1})) + \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2)) \cdot (v(w'_{i+1}, l_i) - v(w'_i, l_i))$
$> \Pr(w'_i \notin l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2)) \cdot (\alpha^d - \alpha^{d-1}) - \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \notin D_2) \cdot (\alpha - 1) + \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2)) \cdot (\alpha^d - \alpha)$
By the statement of the lemma, this quantity is greater than 0. ■

Let us compare these results to what we would get in the preference model of the previous section with additive discounting and multiplicative discounting. The following lemma gives a sufficient and necessary condition for the $(s_1^{\uparrow}, s_2^{\uparrow})$ to be a strict Bayesian-Nash equilibrium for the second level of the ESP game for match-early preferences with additive discounting.

Lemma 22. *Playing words in order of increasing frequency is a strict Bayesian-Nash equilibrium of the second level of the ESP game for any valuation that satisfies additive discounting for match-early preferences if and only if $\Pr(w_i \notin l_{\leq k}(D_2) \cap w_j \in l_{\leq k}(D_2)) > \Pr(w_i \in l_{\leq k}(D_2) \cap w_j \notin D_2)$ for all pairs of words $w_i, w_j \in U$ that satisfy $i < j$ (in other words $f(w_i) > f(w_j)$) and $k \leq |U_{e_1}| - j + 1$, given that $e_1 \leq e_2$.*

Proof. Doing a similar case analysis as the proof of Lemma 20, but no longer differentiating which word we match on, we find that: $\sum_{D_2}(u_1(s'_1(D_1), s_2(D_2)) - u_1(s_1(D_1), s_2(D_2)))$
$= \Pr(w'_i \notin l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2))(v(l_i) - v(l_{i+1})) - \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \notin D_2)(v(l_i) - v(l_{i+1}))$

$\geq \Pr(w'_i \notin l_{\leq i}(D_2) \cap w'_{i+1} \in l_{\leq i}(D_2)) \cdot \alpha - \Pr(w'_i \in l_{\leq i}(D_2) \cap w'_{i+1} \notin D_2) \cdot \alpha$

This quantity must be greater than 0 in order to have a strict Bayesian-Nash equilibrium. ■

It should be noted that an exactly identical statement holds for match-early preferences with multiplicative discounting. It is easy to see from these lemmas that there exist distributions for which $(s_1^{\uparrow}, s_2^{\uparrow})$ is a strict Bayesian-Nash equilibrium for the second level of the ESP game for rare-words-first preferences with additive discounting, but not for match-early preferences with additive discounting. We can make an identical statement when the valuations satisfy multiplicative discounting.

We have left to characterize the sufficient and necessary conditions for playing high effort to be a Bayesian-Nash equilibrium. Our results suggest that this possible since we were able to achieve the $(s_1^{\uparrow}, s_2^{\uparrow})$. The equilibrium strategy $(s_1^{\uparrow}, s_2^{\uparrow})$ is a desirable strategy since players are focusing on the lower frequency, harder to get words, rather than the high frequency, easy to get words. Understanding the incentive structure that leads to high effort is important since it is one way the system designer can extend the set of labels for an image.

References

[1] Y. Benkler. Coase's penguin, or linux and the nature of the firm. *Yale Law Journal*, 112, 2002.

[2] T.-H. Chang, C.-J. Ho, and J. Y. jen Hsu. The photoslap game: Play to annotate. In *Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI)*, pages 1966–1967, 2007.

[3] S. Ficici and A. Pfeffer. Modeling how humans reason about others with partial information. In *Seventh International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2008.

[4] C.-J. Ho, T.-H. Chang, and J. Y. jen Hsu. Photoslap: A multi-player online game for semantic annotation. In *Proceedings of the Twenty-Second Conference on Artificial Intelligence (AAAI)*, pages 1359–1364, 2007.

[5] M. Kearns, S. Suri, and N. Montfort. An experimental study of the human coloring problem on human subject networks. *Science*, 313(5788):824–827, August 2006.

[6] L. von Ahn. Games with a purpose. *IEEE Computer*, 39(6):92–94, 2006.

[7] L. von Ahn and L. Dabbish. Labeling images with a computer game. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems (CHI)*, pages 319–326, 2004.

[8] L. von Ahn, S. Ginosar, M. Kedia, R. Liu, and M. Blum. Improving accessibility of the web with a computer game. In *Proceedings of the 2006 Conference on Human Factors in Computing Systems (CHI)*, pages 79–82, 2006.

[9] L. von Ahn, M. Kedia, and M. Blum. Verbosity: a game for collecting common-sense facts. In *Proceedings of the 2006 Conference on Human Factors in Computing Systems (CHI)*, pages 75–78, 2006.

[10] L. von Ahn, R. Liu, and M. Blum. Peekaboom: a game for locating objects in images. In *Proceedings of the 2006 Conference on Human Factors in Computing Systems (CHI)*, pages 55–64, 2006.