A Game-Theoretic Analysis of the ESP Game[☆]

Shaili Jain[a,*], David C. Parkes[b]

[a] *Computer Science Department, Yale University, New Haven, CT 06520*
[b] *SEAS, Harvard University, 33 Oxford St., Cambridge MA 02139*

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 [3], 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 of the ESP game and characterize the equilibrium behavior in our model. Our equilibrium analysis supports the fact that users appear to be coordinating on low effort words. We provide an alternate model of utility and show that equilibrium behavior in this model achieves more desirable outcomes, from the system designer's perspective. We also give sufficient conditions for coordinating on high effort words to be a Bayesian- Nash equilibrium. 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" [2] (GWAP), 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 [10]. Examples of other peer-production systems include Wikipedia, YouTube, question-and-answer forums such as Yahoo! Answers and Naver Knowledge-iN, and Taskcn, a popular Chinese crowdsourcing website.

Work by von Ahn and others has shown the tremendous power that networks of humans possess to solve problems while playing computer games [3, 5–7, 17, 24]. 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 [3]. 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. As of July 2008, at least

[☆] A preliminary version of this paper appeared at the 4th Workshop on Internet and Network Economics (WINE '08).

[*] Corresponding author

200,000 people have played the ESP game and it has collected over 50 million labels [4]. Subsequent work to the ESP game has included Peekaboom [7], a GWAP for locating objects within an image, Phetch [5], for gathering useful descriptions for images on the web, Verbosity [6], for gathering common sense facts, TagATune [24], for gathering tags for music clips, and Matchin [17], for eliciting user preferences. Still in the spirit of human computation, Kearns et al. [23] provide results from a number of behavioral experiments to see how fast distributed networks of humans can solve various graph problems, such as graph coloring. These authors study how network topology and information constraints change the relative difficulty of various graph problems for these distributed networks of human problem solvers.

While there has been incredible progress in yielding successful applications of human computation, there is still little theory at present to guide design. 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 people coordinate on easy words and that the game is less effective in labeling less obvious, harder words. Google seemed to have noticed this behavior and in their version of the ESP game, called the Google Image Labeler, and introduced different scores for different labels depending on the "descriptiveness" of the label. However, Weber et al. [29], show that this differentiation is not strong enough and that the resulting labels for images still tend to have a high percentage of colors, synonyms or generic words. Ho et al. [20] also notice that the set of labels determined from the ESP game for an image, are not very diverse, and develop a three-player version of the ESP game that involves the addition of a "blocker" to type in words that the other two players cannot use to match.

This paper aims to study behavior in the ESP Game through a game-theoretic light. We propose a simple model of the game in which players independently choose an effort level (low or high), which dictates which portion from a universe of words they sample. If a player samples low effort, she samples from the most frequent set of words in the universe, whereas if a player samples high effort, she samples from the entire universe of words. Once players have independently sampled a dictionary (or type), they decide in which order to output their words. We consider two different models of payoffs, namely *match-early* preferences and *rare-words* 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 an ordinal Bayesian-Nash equilibrium for all distributions on word frequencies, with players focusing attention on high-frequency words. More specifically, we show that choosing low effort in conjunction with playing words in order of decreasing frequency is a Bayesian-Nash equilibrium for all utility functions consistent with match-early preferences and all distributions on word frequencies. For the second stage of the game, we show that playing words in order of decreasing frequency is a Bayesian-Nash equilibrium for all distributions of word frequencies and all utility functions consistent with match-early preferences. Moreover, we show that (decreasing, decreasing) is one of the few second-stage strategy profiles with these properties and we determine that the set of strategy profiles that satisfy this property obey an "almost decreasing" property. Conditioned on the second-stage strategy of playing words in order of decreasing frequency, we show that playing low effort is an ordinal Bayesian-Nash equilibrium and that playing low is an ordinal best response to playing high effort. These results generalize to any number of effort levels. We give the formal statement of the main results obtained in the match-early preferences model below:

Theorem 2. 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 and every choice of effort levels e_1, e_2. Moreover, the set of almost decreasing strategy profiles are the only strategy profiles, in which at least one player plays a consistent strategy, that can be an ordinal Bayesian-Nash equilibrium for every distribution over U and every choice of effort levels e_1, e_2.

Theorem 3. $((L, s_1^{\downarrow}), (L, s_2^{\downarrow}))$ is a strict ordinal Bayesian-Nash equilibrium of the complete ESP game under match-early preferences, for every distribution over U, except the uniform distribution. Moreover, $(L, s_1^{\downarrow})$ is a strict ordinal best-response to $(H, s_2^{\downarrow})$ for every distribution over U, except the uniform distribution.

In order to remedy the problem of users coordinating on common words, which occurs when players adopt low effort and decreasing frequency strategies, we turn to the rare-words preferences model. This is a model in which players wish to match on infrequent words before frequent words, we suppose because of appropriately designed incentives, and where the speed with which a match is achieved is no longer a consideration. We show that under this preference model, there is a significant difference in the equilibrium structure in that playing words in order of decreasing frequency is now a strictly dominated strategy and playing words in order of *increasing* frequency is an ex-post Nash equilibrium. This promotes matching on lower frequency words, with the frequency of the word matched upon, for the same pair of dictionaries, under the (increasing, increasing) strategy profile at least as low as the (decreasing, decreasing) strategy profile. Moreover we show that under the rare-words preferences, we are able to identify, for additional structure on the utility model, an equilibrium behavior that shows a useful focusing on lower frequency words. We show that *high effort* is a Bayesian-Nash equilibrium for Zipfian distributions over word frequencies under certain classes of utility functions that satisfy this preference model. This class of utility functions satisfies an *additive discount* property, meaning that the difference in value between successive outcomes is an additive constant. We give the formal statement of the main results obtained in the rare-words preferences model below:

Theorem 4. Second-stage strategy profile $(s_1^{\uparrow}, s_2^{\uparrow})$ is a strict ex-post Nash equilibrium for the second-stage of the ESP game for every distribution over U and every $e_1 = e_2$, under rare-words preferences.

Theorem 5. $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ is a Bayesian-Nash equilibrium of the complete ESP game for Zipfian distributions over U with $s \leq 1$ and any additive utility function that satisfies rare-words preferences and any multiplicative utility function that satisfies rare-words preferences with $r \geq 2$.

In related work, Hsu and colleagues [11, 19] developed a simple game called PhotoSlap, for determining content of images, based on the popular card game Snap. These authors provide 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. Our model of the ESP game appears to require a more intricate analysis due to the fact that we model it as a game of imperfect information rather than a game of perfect information and the action space for our ESP game model is much larger than the action space for the PhotoSlap model.

von Ahn and Dabbish [4] provide a classification of games with a purpose: *output-agreement games*, such as the ESP game, *inversion-problem games*, such as Peekaboom, and *input-agreement games*, such as TagATune. They provide the key elements of each class in order to ensure the intended computation is done and discuss general design paradigms for increasing enjoyment and output quality. Ho and Chen [21] study the verification mechanisms used in various GWAP and classify the verification mechanisms into two classes, the sequential verification mechanism (as used in inversion-problem and input-agreement games) and the simultaneous verification mechanism (as used in output-agreement games), and model games that use these verification mechanisms. These authors model the simultaneous verification mechanism as a one-shot symmetric coordination game for a report of a single word, and need to appeal to a focal point argument to explain why players will coordinate on the most frequent word. Ho and Chen also model sequential verification games as an extensive form game of imperfect information and show that desirable system wide outcome is supported in an equilibrium. Our focus here is on output agreement games, namely the ESP game.

Weber et al. [29] also notice that the ESP Game and Google Image Labeler fail to collect informative labels. In a thoughtful experimental study of the data generated from the Google Image Labeler, Weber et al. [29] show that the set of tags for a given image are generated from a low entropy distribution and that the new labels to be entered by players are highly predictable given the Taboo Words. In order to make the case for the second point, these authors program a bot based on the language model used from Google Image Labeler and the bot infers what label should come next solely from the set of Taboo Words already present and derives no information from the image itself. They find that this bot agrees with a human player 81% of the time on images that have at least one Taboo Word.[1] Weber et al. [29] also determine that over 10% of the Taboo Words are colors. In order to remedy this problem, these authors propose two alternate scoring schemes, such as rewarding players for a label with value inversely proportional to the probability that this label would be entered given the set of Taboo Words and rewarding players based on the amount of information gain from each new label, but do not provide a game-theoretic analysis of these schemes. In contrast to their work, our work seeks to understand under what conditions it is possible to have entering descriptive labels be supported by equilibrium analysis.

There have been game-theoretic models of other peer production systems, including a study of scoring mechanims in Yahoo! Answers [22], all-pay auction models of crowdsourcing systems (such as Taskcn and TopCoder) [8, 12, 14] and related work in economics on the optimal design of *contests* [26, 27], and analysis of attention mechanisms in social computing systems [15, 16]. In addition to this, there have been a number of empirical studies of user behavior in various peer production systems which show that some fraction of users in these systems are behaving strategically [1, 28, 30], motivating the use of game theory to study such systems.

2. The ESP Game

The ESP Game [3] 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 [9].

[1] This bot is programmed to behave more "human-like", by entering words at a human-like speed and entering a maximum of 10 words per image.

However, in order to label images, humans require some sort of incentive for this normally tedious task. This is achieved in the ESP game by making the game fun to play.

In the ESP game, 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 for each successful label. In the set of 15 images, players get bonus points labeling five images, ten images, and fifteen images. Players can pass on difficult images and they are revisited at the end of a set. The only word that is used from the two input streams of an image is the first common word that is entered. 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. The game includes a scoreboard, with the names of players with the highest scores, that is updated daily. Empirical studies of other peer-production systems has shown that points are a key feature in motivating users [28].

2.1. A Formal ESP Model

We model the ESP game as a two-stage game of imperfect information. In our model, when a player decides to play the ESP game, she is presented with an image and thinks up words to represent the image. She then makes a decision about how to enter words depending how likely she is to match with the other player on those words. We focus on modeling the game associated with one of the images in a set.

Let there be a universe of n words $U = \{w_1, w_2, ..., w_n\}$ associated with the image at hand and let $1 < d < n$ denote the *dictionary size*, or the number of words that each player samples from the universe[2]. Each word in the universe has an associated frequency, where f_i denotes the frequency of word w_i in the English language and $\sum_{i=1}^{n} f_i = 1$. Each player knows the frequency of the words sampled and can therefore rank words according to frequency[3].

Even though this is a game without any communication between players, it is 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*: $E = \{L, H\}$ for *low* 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 $n > n_L > 0$ words (without replacement). That is, word i in the top n_L words is chosen first with probability $f_{i,L} = \frac{f_i}{\sum_{j=1}^{n_L} f_j}$. Let U_L be the set of the highest n_L frequency words in U, or the "low universe". In addition, let $\mathcal{D}_L$ denote the set of all possible dictionaries a player could obtain if she played L effort. If a player chooses effort H, the dictionary is sampled from the entire universe, without replacement. That is word i in U is chosen with probability $f_{i,H} = f_i$. Similarly, $\mathcal{D}_H$ denotes the set of all possible dictionaries a player could obtain if she played H effort. Note that we assume $d < n_L$.

Given a word $x \in U$, we let $f_e(x)$ represent the probability of sampling x given that the player has chosen effort level e. This sample is modeled as a move by nature and can be considered to

[2]Sometimes we use the additional assumption that $d \leq \frac{n}{2}$.

[3]Additionally, assume that the words in the universe are ordered in terms of decreasing frequency, that is $f_1 \geq f_2 \geq ... \geq f_n$.

be the point at which a player learns her "type", namely her dictionary of words. Both players are symmetric and each player has the same decision space. Note that n_L, n_H, and d are parameters of the model.

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 as the maximum value of the two positions where the word occurs in each ordered list) at which that occurs.

Let D_1 be the dictionary for player 1 and D_2 be 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.

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, which 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 is a match in location $max(i, j)$. The *first* match 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. Let *outcome function* $g(s_1(D_1), s_2(D_2)) \in \mathcal{O}$ denote the outcome given s_1, s_2, D_1, and D_2, with the location (if any) of the first match is denoted $g_l(s_1(D_1), s_2(D_2)) \in \mathcal{L}$ and the word the first match occurs is denoted $g_w(s_1(D_1), s_2(D_2)) \in U$.

Each player i has a utility function $v_i : \mathcal{O} \rightarrow \Re^+$ which induces a weak total preference ordering on outcomes. We assume that both players have the same utility function. We consider two preference models: *match-early preferences* and *rare-words preferences*. In both cases, we work with an *ordinal* model of preferences.

Definition 2. For match-early preferences, we require the following preference ordering on outcomes: $(w_1, l_1) \equiv (w_2, l_1) \equiv ... \equiv (w_n, l_1) \succ (w_1, l_2) \equiv (w_2, l_2) \equiv ... \equiv (w_n, l_2) \succ ... \succ (w_1, l_d) \equiv (w_2, l_d) \equiv ... \equiv (w_n, l_d) \succ (\phi, \phi)$ for all players.

Since players are indifferent between which word they match upon under match-early preferences, we can simply describe the outcome of the match as a location, i.e. l_i can be used to describe any element in the set $\{(w_1, l_i), (w_2, l_i), ..., (w_n, l_i)\}$. We say that a utility v_i is consistent with match-early preferences if and only if $v_i(l_1) > v_i(l_2) > ... > v_i(l_d) > v_i(\phi)$. This preference model captures the fact that 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.

Definition 3. For rare-words preferences, we require the following preference ordering on outcomes: $(w_n, l_1) \equiv (w_n, l_2) \equiv ... \equiv (w_n, l_n) \succ (w_{n-1}, l_1) \equiv (w_{n-1}, l_2) \equiv ... \equiv (w_{n-1}, l_n) \succ ... \succ (w_1, l_1) \equiv (w_1, l_2) \equiv ... \equiv (w_1, l_n) \succ (\phi, \phi)$.

Under rare-words preferences, players are indifferent between which location they match and only care about which word they match upon. Therefore we can simply use a word to denote the outcome, i.e. w_i can be used to describe any element in the set $\{(l_1, w_i), (l_2, w_i), ..., (l_d, w_i)\}$. We say that a utility function is consistent with rare-words preferences if and only if $v_i(w_n) > v_i(w_{n-1}) > ... > v_i(w_1) > v_i(\phi)$.

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 4. 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 $p(l_i, s_1(D_1), s_2) = \sum_{D_2} \Pr(D_2) I(g_l(s_1(D_1), s_2(D_2)) = l_i)$.

Definition 5. The probability of first match on word w_j given $s_1(D_1)$, s_2, and a distribution on dictionaries $\Pr(D_2)$ is $p(w_j, s_1(D_1), s_2) = \sum_{D_2} \Pr(D_2) I(g_w(s_1(D_1), s_2(D_2)) = w_j)$. Often times we will abbreviate $p(l_i, s_1(D_1), s_2)$ as $p(l_i)$ and $p(w_j, s_1(D_1), s_2)$ as $p(w_j)$.

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. Another way to express the expected interim utility to player i given dictionary D_i uses the probability of first match vector defined above: $u_i(s_i(D_i), s_{-i}) = \sum_{l \in \mathcal{L}} p(l, s_i(D_i), s_{-i}) \cdot v_i(l)$. 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 (ex-ante) utility to player i before either dictionaries are sampled, given complete strategies $\sigma = (\sigma_1, \sigma_2)$.

2.2. Equilibrium Framework

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 a player might still usefully deviate to an alternate effort in combination with an alternate second stage strategy. To continue, consider the game induced by fixing 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 some useful equilibrium concepts:

Definition 6. 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 equilibrium is strict as long as there exists a pair of D_1, D_2 such that the above inequality is strict.

We will adopt an analysis approach that establishes a strict *ordinal* Bayesian-Nash equilibrium[4], in the sense that we identify strategies that are an equilibrium for all utility functions consistent with match-early preferences.

Definition 7. 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 every D_i, and every u_i consistent with match-early preferences,

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

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.

There is an identical definition of ordinal Bayesian-Nash equilibrium for the second-stage of the ESP game for rare-words preferences. We also define ordinal Bayesian-Nash equilibrium for the entire game.

Definition 8. A strategy profile $\sigma^* = (\sigma_1^*, \sigma_2^*) \in \Sigma_1 \times \Sigma_2$ is a strict ordinal Bayesian-Nash equilibrium of the ESP game if for every u_i consistent with match early preferences, we have

$$u_i(\sigma_i^*, \sigma_{-i}^*) > u_i(\sigma_i', \sigma_{-i}^*)\ \forall \sigma_i' \neq \sigma_i^*, \quad \forall i \in \{1,2\} \tag{3}$$

Likewise, there is an identical definition of ordinal Bayesian-Nash equilibrium for the complete ESP game for rare-words preferences. Since the effort level chosen by each player is not visible to the other player, there is no need for a subgame perfect refinement.

Next we define the notion of stochastic dominance for a general utility function. Our definition uses the following notation: Suppose that $u(s_i, s_{-i}) = \sum_{o \in \mathcal{O}} \Pr(o|s_i, s_{-i}) \cdot v_i(o) = \sum_{k=1}^{m} \Pr(o_k|s_i, s_{-i}) \cdot v_i(o_k)$, for some ordering $o_1, o_2, ..., o_m$ on outcomes. We say that the strategies s_i and s_{-i} induce a probability vector on outcomes $(\Pr(o_1|s_i, s_{-i}), \Pr(o_2|s_i, s_{-i}), ..., \Pr(o_m|s_i, s_{-i}))$.

Definition 9. Strategy s_i stochastically dominates s_i' with respect to opponent strategy s_{-i} and outcome ordering $o_1, o_2, ..., o_m$ if and only if $\sum_{a=1}^{k} \Pr(o_a|s_i, s_{-i}) \geq \sum_{a=1}^{k} \Pr(o_a|s_i', s_{-i})$ for all k. We say that the stochastic dominance property is strict if there exists a k such that $\sum_{a=1}^{k} \Pr(o_a|s_i, s_{-i}) > \sum_{a=1}^{k} \Pr(o_a|s_i', s_{-i})$.

The following theorem equates our definition of stochastic dominance and ordinal Bayesian-Nash equilibrium. We omit the proof of the following theorem since it is a standard proof in stochastic dominance [18]. The literature on ordinal Bayesian incentive-compatibility for representation of committees, stable matchings, etc. likewise uses an analogous definition of stochastic dominance to establish that truth-telling is an ordinal Bayesian-Nash equilibrium [13, 25].

[4]It should be noted that typically, the Bayesian-Nash equilibrium concept is used in games of incomplete information rather than games of imperfect information. In this particular game, when considering Bayesian-Nash equilibrium, the utility is computed in expectation over the distribution of choices by nature (i.e. distribution over all possible dictionaries). In incomplete information games, when considering Bayesian-Nash equilibrium, the utility is computed in expectation over the distribution over agent types (i.e. distribution over agent's private valuation). In our model, agents' valuations are assumed to be common knowledge, thus our game is not an incomplete information game.

Theorem 1. *Strategy s_i strictly stochastically dominates s'_i with respect to opponent strategy s_{-i} if and only if $u_i(s_i, s_{-i}) > u_i(s'_i, s_{-i})$ for all v_i consistent with the preference ordering $o_1 \succ o_2 \succ ... \succ o_m$.*

This means that we can use the stochastic dominance condition to establish ordinal Bayesian-Nash equilibrium for the second-stage of the ESP game and the complete ESP game. We define stochastic dominance more specifically for the second-stage game and the complete ESP game under each preference model as needed.

2.3. Remarks about the Model

We model the ESP game with each player sampling words from a universe of possible words associated with the image, 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. We capture the idea that there are 15 images in a set with a limited amount of time by considering preferences in which an agent prefers to match in an earlier location rather than a later location. We do not model a sequential decision making process, where users choose an effort level before sampling and entering each successive word. We omit this because there seems to be little inference a player can make about the strategy of the other player from the limited information revealed; all a player learns is that no match has occurred, assuming that players are entering words at roughly the same speed. Given the random sampling of words from the universe this provides little evidence to discriminate between a player playing her words in order of decreasing frequency and a player playing her words in order of increasing frequency. Rather it seems more likely that strategy updates occur after a successful match, where it is learnt what word has been used to match. In addition, it seems as though the time frame per image is rather small, e.g. 2.5 minutes for 15 images. Thus it seems as though users enter a small number of words per image and move on to the next image, without updating their strategy in between. It would be interesting to empirically analyze the data from the ESP game to examine where strategy updates occur, among other things.

The universe is the set of words that are in some way relevant to the image and represent the knowledge that the game designer is trying to learn. We assume that each of these words is relevant and accurate with respect to the image at hand. For example, if we had an image of a Victorian house and we had the two labels, "building" and "Victorian house", both are relevant and accurate, yet one is far more descriptive than the other.

As mentioned above, we decompose the game into a "first stage", or choice of effort level, and a "second stage", or choice of permutation on a sampled dictionary. Note that the use of "stage" here is an abuse of notation since there is no observable action after the choice of effort level. It is more appropriate to refer to what we call the "second stage" as the *effort-constrained game*. This is the game induced by a pair of effort levels chosen by the players. We often use "second stage" and "effort-constrained game" interchangeably. Given the equilibrium analysis in the effort-constrained game, we can analyze the complete game, by fixing the strategies determined in the effort-constrained game analysis and examining the choice of effort level (or first stage decision). We refer to this analysis as the complete game analysis. It is important to note that all the results in the paper generalize to any number of effort levels (up to n effort levels), but we describe all the results using two effort levels for simplicity.

In sampling dictionaries, we believe it is reasonable to assume that players sample their words for any given image according to the distribution induced by the frequencies of words in the English language because there is cognitive effort required to retrieve less frequently used words. Likewise, the English language is coded efficiently in that the more frequent, common words are generally shorter whereas the less frequent, more descriptive words are generally longer. 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.

Match-early preferences are meant to capture that players prefer to match sooner rather than later due to the time-constraint in the ESP game. The actual implementation of the ESP game assigns the same number of points to players if they match, regardless of where the match occurs (e.g. how many words they enter before they match) and regardless of which word the match occurs. Despite this, it is obvious that players are under a time constraint and they prefer to match sooner rather than later, in order to achieve a match on as many images as possible in the allotted amount of time. We work with an ordinal model of preferences so that we do not have to quantify exactly how much players prefer to match sooner rather than later. Rare-words preferences are meant to capture that players prefer to match on rarer words than more frequent words and are indifferent between the location that they match on, presumably because users will be given more time for a set of images such that time is no longer a key constraint.

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 of the second-stage of the ESP game 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 second-stage strategy involves specifying a total ordering of elements on U_e (after choosing an effort level e) and applying that total ordering to the realized dictionary. 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. A consistent strategy s specifies a total ordering on the set U_e: $w'_1 \succ w'_2 \succ ... \succ w'_{|U_e|}$, 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 her words in order of decreasing frequency. We believe that equilibrium in consistent strategies are natural because of their simplicity, requiring that word x is played before word y independent of whether any word z is present in a player's dictionary.

3. Equilibrium Analysis 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 an ordinal Bayesian-Nash equilibrium for the ESP game.

3.1. Equilibrium Analysis of the Effort-Constrained Game

First we see that playing words in order of decreasing frequency is not an ex-post Nash equilibrium for the second stage game.

Proposition 1. *Suppose that* $e = \min(e_1, e_2)$*, there are three words in the universe* U_e*,* $\{w_1, w_2, w_3\}$*, and* $d = 2$*. The second-stage strategy profile* $s = (s_1^{\downarrow}, s_2^{\downarrow})$ *is not an ex-post Nash equilibrium.*

Proof. Suppose $D_2 = \{w_2, w_3\}$ and $D_1 = \{w_1, w_2\}$, $s_2(D_2)$ dictates player 2 will play w_2 followed by w_3. If player 1 deviates from s_1 and plays w_2 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 establishing ordinal Bayesian-Nash equilibrium via stochastic dominance. We define stochastic dominance for the ordering on outcomes under match-early preferences, i.e. with matching in the first location as the most preferable outcome, matching in the second location as the second most preferable outcome, etc.

Definition 10. 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 $s_1(D_1)$ *stochastically dominates* the second-level ordering $s'_1(D_1)$ with respect to match-early preferences 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 10 gives 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 $D_1 \in \mathcal{D}_1$.

In what follows, we show that "playing decreasing frequency" is a strict ordinal Bayesian-Nash equilibrium of the second-stage ESP game, for any pairs of effort levels e_1, e_2 and for any distribution over U. Moreover, we show that this equilibrium is one of the few ordinal Bayesian-Nash equilibrium that holds for every distribution over U. The set of strategy profiles that are ordinal Bayesian-Nash equilibrium of the second-stage game satisfy an "almost decreasing" property. The crux of the argument will be to establish stochastic dominance.

Below we describe a possible 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, 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.

Algorithm 1 Candidate Best Response for Player 1

Input: sampled D_1, $\sigma_2 = (e_2, s_2)$
Maintain ordered list $s_1(D_1) = \emptyset$
for $i = 1$ to d **do**
  Add element

$$E_{add} = \operatorname*{arg\,max}_{w_j \in D_1 - s_1(D_1)} \sum_{D_2 \in \mathcal{D}_{e_2}} \Pr(D_2) \cdot I(w_j \text{ is in the top } i \text{ of } s_2(D_2))$$

  to the end of the ordered list $s_1(D_1)$
end for
Output: $s_1(D_1)$

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 computes a feasible output ordering 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 *in agreement* 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. Recall that we look for equilibrium in consistent strategies and so s_2 is associated with a well-defined ordering.

This algorithm does not always output an ordering that stochastically dominates all other orderings in the sense of Definition 10. But, any time it fails to produce such an output, we show that no such ordering exists.

The following definition is useful in characterizing the output of Algorithm 1. Note that the set $\{w'_1, ..., w'_n\}$ is ordered according to s_2, i.e. s_2 specifies playing the following total order on words: $w'_1 \succ w'_2 \succ ... \succ w'_n$. 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 11. 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 $\max(1, i - n + d) \leq k \leq \max(i, d - 1)$.

Definition 12. We say that s_2 satisfies the *strong 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 $i < j - 1$, we have that $\Pr(w'_i \in D_2) > \Pr(w'_j \in l_{\leq k}(s_2(D_2)))$ for all $\max(2, i - n + d + 1) \leq k \leq \min(j - 1, d - 1)$.

In an almost decreasing strategy the first $n - 1$ words of s_2 are sorted in order of the decreasing frequency, but the last word may not necessarily be the least frequent word of U. Therefore, there are a total of n strategies that satisfy this property. An *almost decreasing strategy profile* to describe a symmetric strategy profile (s, s), where s is an almost decreasing strategy.

Definition 13. We say that a consistent strategy s_2 satisfies the *almost decreasing* property if and only if $f(w'_i) > f(w'_j)$, for all $1 \leq i < j \leq n - 1$.

Lemma 1. *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. Let the output of Algorithm 1 be $s(D_1)$. Suppose there exists another ordering $s'(D_1) \neq s(D_1)$ that stochastically dominates all other strategies. Let l_i be the first coordinate in which $p(s(D_1), s_2)$ and $p(s'(D_1), s_2)$ differ. 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. Therefore, $\sum_{a=1}^{i} p(l_a, s(D_1), s_2) \geq \sum_{a=1}^{i} p(l_a, s'(D_1), s_2)$. $\square$

Recall from Theorem 1, that stochastic dominance, as defined in Definition 10, is a necessary condition in order to have utility maximization for all utilities consistent with match early preferences.

The following lemma gives sufficient conditions on the strategy of player 2 such that Algorithm 1 will always output an ordering in agreement with s_2, and such that this strategy will stochastically dominate all other strategies.

Lemma 2. *If second-stage strategy s_2 satisfies the preservation condition for a particular distribution, then Algorithm 1 will always output an ordering in agreement 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 in agreement with s_2 will strictly stochastically dominate all other strategies, for any sampled dictionary.*

Proof. Since s_2 satisfies the preservation condition, for every $D_1 = \{w''_1, w''_2, ..., w''_d\}$ with $w''_1 \succ w''_2 \succ ... \succ w''_d$ under s_2, $\Pr(w''_i \in l_{\leq k}(s_2(D_2))) > \Pr(w''_j \in l_{\leq k}(s_2(D_2)))$ for all $1 \leq k \leq i$. Since $\Pr(w''_i \in l_{\leq k}(s_2(D_2))) > \Pr(w''_j \in l_{\leq k}(s_2(D_2)))$ for all $1 \leq k \leq i$, w''_j cannot be output before w''_i by Algorithm 1, for every D. Since this is true for all $w''_i, w''_j \in D$ with $w''_i \succ w''_j$, Algorithm 1 must output an ordering in agreement with s_2. The strong condition tells us that, for every D_1, where $D_1 = \{w''_1, w''_2, ..., w''_d\}$ with $w''_1 \succ w''_2 \succ ... \succ w''_d$ under s_2, $\Pr(w''_i \in D) > \Pr(w''_j \in D(k))$ for all i, j, k, with $i < k < j$. Thus w''_1, w''_2, ..., w''_i are the set of words that strictly maximize $\sum_{j=1}^{i} p(l_j, s_1(D_1), s_2)$ for any i, and stochastic dominance is satisfied.

□

Lemma 3. *If a consistent strategy s_2 satisfies the almost decreasing property, then the strategy profile (s_2, s_2) is a strict ordinal Bayesian-Nash equilibrium of the second-stage ESP game under match-early preferences, for every choice of effort levels e_1 and e_2, for every distribution over U.*

Lemma 4. *The symmetric strategy profile (s_2, s_2) is a strict ordinal Bayesian-Nash equilibrium of the second-stage ESP game under match-early preferences, for every s_2, for every $e_1 = e_2$, for the uniform distribution over U.*

Proof. Since the distribution over U is uniform, for all w'_i, w'_j with $w'_i \succ w'_j$ in s_2, we have $\Pr(w'_i \in D_2) = \Pr(w'_j \in D_2)$. This gives us $\Pr(w'_i \in D_2 \cap w'_j \notin D_2) = \Pr(w'_j \in D_2 \cap w'_i \notin D_2)$. 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*, for every s_2. From Lemma 2 we get that s_2 is a best response to s_2. Finally, 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 2, along with Theorem 1, gives us the desired result. □

This lemma tells us that that strategy profile (s'_1, s_2), with $s'_1 \neq s_2$, cannot be an ordinal Bayesian-Nash equilibrium under the uniform distribution, because under the uniform distribution over U, for every sampled dictionary, s_2 generates a distribution on outcomes that strictly stochastically dominates $s'_1 \neq s_2$. Therefore, for every utility function consistent with match-early preferences, player 1 would prefer to deviate from s'_1 to s_2. Therefore, if a strategy profile is an ordinal Bayesian-Nash equilibrium for *all* distributions over U, then it must be that case that the strategy profile is symmetric.

It should be noted that the statement of Lemma 4 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 s_2, e.g. all words in $U_{e_1} - U_{e_2}$ are lower priority than all words in U_{e_2} under s'_2.

Lemma 5 tells us that for every symmetric strategy profile (except for the almost decreasing strategy profiles), there exists a distribution such that this strategy profile is not an ordinal Bayesian-Nash equilibrium. Lemma 4 rules out the possibility of an asymmetric strategy profile as an ordinal Bayesian-Nash equilibrium that holds for every distribution over U and Lemma 5 rules out the possibility of a symmetric strategy profile (except for the almost decreasing strategy profiles) as an ordinal Bayesian-Nash equilibrium that holds for every distribution over U. Therefore, Lemmas 4 and 5 can be used to establish Theorem 2, where it is shown that the almost decreasing strategy profiles are the only strategy profiles that are an ordinal Bayesian-Nash equilibrium for every distribution over U.

Lemma 5. *For every (e_2, s_2) (except for s_2 that are almost decreasing), there exists a distribution over U, an effort level e_1, and a dictionary for which Algorithm 1 will not output an ordering in agreement with s_2.*

Theorem 2. *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 and every choice of effort levels e_1, e_2. Moreover, the set of almost decreasing strategy profiles are the only strategy profiles, in which at least one player plays a consistent strategy, that can be an ordinal Bayesian-Nash equilibrium for every distribution over U and every choice of effort levels e_1, e_2.*

Proof. Lemma 3 tells us that Algorithm 1 will always output a strategy in agreement with $s_2^{\downarrow}$, if player 2 is playing $s_2^{\downarrow}$, regardless of D_1 and the distribution over U. Furthermore, this strategy stochastically dominates all other strategies, for every distribution over U. Lemma 4 tells us that there exists a distribution, namely the uniform distribution, for which Algorithm 1 will output an ordering in agreement with s_2, regardless of the dictionary D_1, for all s_2 that are consistent. Moreover, this strategy will stochastically dominate all others, for the uniform distribution. Lemma 5 tells us that there exists a distribution $F(U)$ and dictionary D_1 for which Algorithm 1 will output an ordering that is not in agreement with s_2, for all s_2 that are not almost decreasing. 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 1. Therefore, there is no single strategy for player 1 that stochastically dominates all others when player 2 is playing s_2, where s_2 is not almost decreasing, for all distributions over U and every utility function that satisfies match-early preferences. □

Definition 14. We say that the distribution on words in the universe satisfies a Zipfian distribution if and only if $f(w_i) = \frac{1}{i^s}$ for any $s > 0$. We restrict attention to the case where $s \geq 1$.

Lemma 6. *If there exists an $1 \leq i \leq n$ such that $\Pr(w'_i \notin l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_{i+1} \in l_{\leq j}(s_2^{\uparrow}(D_2))) > \Pr(w'_i \in l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_{i+1} \notin l_{\leq j}(s_2^{\uparrow}(D_2)))$ for some $\max(1, i-n+d) \leq j \leq \max(i, d-1)$, then $(s_1^{\uparrow}, s_2^{\uparrow})$ cannot be a Bayesian-Nash equilibrium for any utility function satisfying match-early preferences.*

Proposition 2. *Second-stage strategy profile $(s_1^{\uparrow}, s_2^{\uparrow})$ cannot be a Bayesian-Nash equilibrium for the second-stage of the ESP game for any Zipfian distribution over U with $s \geq 1$ and for any utility function satisfying match-early preferences.*

3.2. Equilibrium Analysis 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 a strict ordinal Bayesian-Nash equilibrium for all distributions except the uniform distribution over U. For the case of the uniform distribution, $((L, s_1^{\downarrow}), (L, s_2^{\downarrow}))$ is a weak ordinal Bayesian-Nash equilibrium. In order to show this, we first carefully specify what it means for a strategy to stochastically dominate another for the top level of the game, which fixes the equilibrium strategy for the bottom-level. This 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 15. 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, with respect to the outcome ordering of match-early preferences, if and only if:

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

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 3 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 $I(g_l(s_1(D_{1,e_1})(k), s_2(D_{2,e_2})(k)) \in \{l_1, ..., l_k\}) = I(D_{1,e_1}(k) \cap D_{2,e_2}(k) \neq \emptyset)$ 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.

In order to establish stochastic dominance, we construct a randomized mapping for each dictionary that can be sampled when playing H to a number of dictionaries that can be sampled when playing L. Each dictionary in $\mathcal{D}_H$ 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 8. 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 9.

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 needed to prove Lemma 8.

Lemma 7. *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 $D' = \{w'_1, w'_2, ..., w'_n\}$ and $D = \{w_1, w_2, ..., w_n\}$ (both in sorted order) where $w_j = w'_j$ for all $j \neq i$ and $f(w'_i) > f(w_i)$. 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 D satisfies property P if $D(k) \cap A = \emptyset$ and $w_i \in D(k)$ and property P' if $D(k) \cap A = \emptyset$ and $w'_i \in D(k)$. It suffices to show that $\Pr(D_2 \text{ satisfies } P') > \Pr(D_2 \text{ satisfies } P)$. Consider the transformation $t : D_i \to D'_i$, where D_i is any dictionary that satisfies P. If D_i also has w'_i, t maps D_i to D_i and note D_i satisfies property P'. If D_i does not have w'_i, t 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 17, $\Pr(D'_i) > \Pr(D_i)$. 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 $D \in \mathcal{D}_H$, $D = A \cup B$, where A is the set of "low words" and B is the set of "high words" (in other words, $A = D \cap U_L$ and $B = D \cap (U_H - U_L)$). 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_H (without replacement) until you got d "low words".

Note that if D contains only high 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 8. *For every $D_{1,H}$, where $D_{1,H}$ is a dictionary sampled with respect to the H effort level, and for every h that satisfies the property that $D_{1,H}$ is mapped to a dictionary in $\mathcal{D}_L$ that contains the set $D_{1,H} \cap U_L$ and every effort level of player 2 and when both players play decreasing frequency in the second stage, we have that:*

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

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

Proof. Due to Lemma 7, it suffices to show $h(D_{1,H}) = \{w'_1, w'_2, ..., w'_d\}$ dominates $D_{1,H} = \{w_1, w_2, ..., w_d\}$. Assume there exists a coordinate i with $f(w'_i) < f(w_i)$. Let j be the minimum such coordinate. $w_j \in U_L$ since $h(D_{1,H})$ contains words only in U_L. This means $w_j \in h(D_{1,H})$. Since the dictionaries are in sorted order, this means $w_j = w'_k$ for some $k < j$, however, this means $h(D_{1,H})$ does not contain all of $D_{1,H} \cap U_L$, a contradiction. When $h(D_{1,H}) \neq D_{1,H}$, the dominance is strict. □

Lemma 9 states that the distribution obtained from sampling U_L directly is the same as the distribution obtained from sampling a high dictionary, followed by the randomized mapping (i.e. sampling U_H until you get d low words). The proof is easy and omitted.

Lemma 9. $\Pr(D_{1,L}|L) = \sum_{D_{1,H}} \Pr(D_{1,H}|H) \cdot \Pr(h(D_{1,H}) = D_{1,L})$

Lemma 10 uses Lemmas 8 and 9 to show that playing L stochastically dominates playing H under match-early preferences, 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 continues to hold as we vary the number of effort levels, as long as there are at least two.

Lemma 10. *For every effort level of player 2 and when players play decreasing frequency in the second stage:*

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

Proof. From Lemma 9, we know that:

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

This is equivalent to writing:

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

where the inequality follows from Lemma 8. We know that the inequality is strict for all k since there exists a $D_{1,H}$ such that $h(D_{1,H}) \neq D_{1,H}$ and $h(D_{1,H})$ and $D_{1,H}$ differ in the first coordinate, for all possible values of $h(D_{1,H})$ under the randomized mapping. □

Theorem 3 together with Lemma 10 and Theorem 1 gives us the following result.

Theorem 3. $((L, s_1^{\downarrow}), (L, s_2^{\downarrow}))$ *is a strict ordinal Bayesian-Nash equilibrium of the complete ESP game under match-early preferences, for every distribution over U, except the uniform distribution. Moreover,* $(L, s_1^{\downarrow})$ *is a strict ordinal best-response to* $(H, s_2^{\downarrow})$ *for every distribution over U, except the uniform distribution.*

Corollary 1. $((H, s_1^{\downarrow}), (H, s_2^{\downarrow}))$ *is not a Bayesian-Nash equilibrium of the complete ESP game for any distribution over U, except the uniform distribution, and for any utility function that satisfies match-early preferences.*

4. The Effect of Rare-Words Preferences

In this section, we consider the effect of rare-words preferences on the equilibrium analysis. We look to understand whether there is a *high effort* equilibrium available in this preference model, when players care about matching on rare words and are indifferent between the location that they match.

For the second stage, we show that playing words in order of *decreasing frequency* is strictly dominated, in stark contrast with the previous section. Also, playing words in order of *increasing frequency* is an ex-post Nash equilibrium of the second-stage game, for all pairs of effort levels chosen in the first stage. Although, we show that $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ cannot be an ordinal Bayesian-Nash equilibrium in the complete game for any distribution over U, we show that for every distribution over U, there exists a utility function for which $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ is a Bayesian-Nash equilibrium. We also show that $((L, s_1^{\uparrow}), (L, s_2^{\uparrow}))$ is an ordinal Bayesian-Nash equilibrium for every distribution over U and this leads to a better outcome from the system designer's perspective for every pair of player dictionaries. Finally, we demonstrate sufficient conditions on the utility function for Zipfian distributions in order for $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ to be a Bayesian-Nash equilibrium under rare-words preferences.

4.1. Equilibrium Analysis of the Effort-Constrained Game

We show that in this model, playing words in order of increasing frequency is not a dominant strategy equilibrium.

Definition 16. Second-stage strategy profile s_1^* is a dominant strategy 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:

$$u_1(s_1^*(D_1), s_2(D_2)) \geq u_1(s_1'(D_1), s_2(D_2)),\ \forall s_2 \text{ and } s_1' \neq s_1^* \tag{7}$$

Definition 17. Second-stage strategy profile s_1 is a dominated strategy 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, there exists an s_1' such that:

$$u_1(s_1'(D_1), s_2(D_2)) \geq u_1(s_1(D_1), s_2(D_2)),\ \forall s_2 \tag{8}$$

We say that s_1 is strictly dominated if there also exists a D_1, D_2 and s_2 such that the above inequality is strict.

Proposition 3. *Suppose that* $e = \min(e_1, e_2)$*, there are five words in the universe* U_e*,* $\{w_1, w_2, w_3, w_4, w_5\}$*, and* $d = 4$*. The second stage strategy profile* $s = (s_1^{\uparrow}, s_2^{\uparrow})$ *is not a dominant strategy equilibrium in the second-stage game for any distribution over* U *under rare-words preferences.*

Proof. Suppose $D_2 = \{w_1, w_2, w_4, w_5\}$, $s_2(D_2) \to w_4 \succ w_2 \succ w_5 \succ w_1$, and $D_1 = \{w_1, w_3, w_4, w_5\}$. If $s_1 = s_1^{\uparrow}$, the match will occur on w_4. However, if $s_1 \to w_5 \succ w_3 \succ w_1 \succ w_4$, the match will occur on w_5, a strictly better outcome. □

While playing words in order of increasing frequency is not a dominant strategy equilibrium of the second-stage game, it is an ex-post Nash equilibrium with rare-words preferences.

Theorem 4. *Second-stage strategy profile $(s_1^{\uparrow}, s_2^{\uparrow})$ is a strict ex-post Nash equilibrium for the second-stage of the ESP game for every distribution over U and every $e_1 = e_2$, under rare-words preferences.*

The statement of the above theorem can be generalized to handle the case where players are playing different effort levels. In the case that player 1 is playing a higher effort level than player 2, player 1's best response to is play increasing on the set $D_1 \cap U_{e_2}$, followed by the words in the set $D_1 \cap (U_{e_1} - U_{e_2})$ (in any order). Likewise, in the case that player 1 is playing a higher effort level than player 1, player 2's best response is to play increasing on the set $D_2 \cap U_{e_1}$ (in any order), followed by the words in the set $D_2 \cap (U_{e_1} - U_{e_2})$ (in any order). This "generalized" increasing strategy can be shown to be an ex-post Nash equilibrium.

Proposition 4. *Second-stage strategy $s_1^{\downarrow}$ is strictly dominated for any second-stage strategy of player 2 and for any distribution over U and any choice of effort levels e_1, e_2, under rare-words preferences.*

4.2. Equilibrium Analysis of the Complete Game

In order to analyze the top-level game, we define stochastic dominance under the order on outcomes associated with rare-words preferences.

Definition 18. 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 with respect to the ordering on outcomes given by rare-words preferences, if and only if:

$$\sum_{D_{1,e_1}} \Pr(D_{1,e_1}|e_1) \sum_{D_{2,e_2}} \Pr(D_{2,e_2}|e_2) I(g_w(s_1(D_{1,e_1}), s_2(D_{2,e_2})) \in \{w_n, ..., w_{n-k+1}\}) \geq$$
$$\sum_{D_{1,e_1'}} \Pr(D_{1,e_1'}|e_1') \sum_{D_{2,e_2}} \Pr(D_{2,e_2}|e_2) I(g_w(s_1(D_{1,e_1'}), s_2(D_{2,e_2})) \in \{w_n, ..., w_{n-k+1}\}) \ \forall k$$

In addition, we say the stochastic dominance is *strict* if there exists a k such that the above inequality is strict.

The next two propositions give general results about strategies in the complete ESP game and use the randomized mapping from the previous section to map each high dictionary to a low dictionary that does at least as well as it in expectation. A designer will prefer this equilibrium to the $((L, s_1^{\downarrow}), (L, s_2^{\downarrow}))$ equilibrium we found for match-early preferences, since the $((L, s_1^{\uparrow}), (L, s_2^{\uparrow}))$ equilibrium leads to matches on a "rarer" word. We formalize this observation in Remark 1.

Proposition 5. $((L, s_1^{\uparrow}), (L, s_2^{\uparrow}))$ *is a strict ordinal Bayesian-Nash equilibrium of the complete ESP game for every distribution over U under rare-words preferences.*

Proof. We use the randomized mapping from section 4 to map each $D_H \in \mathcal{D}_H$ to a set of dictionaries in $\mathcal{D}_L$, such that D_H is mapped to D_L if and only if $D_H \cap U_L \subseteq D_L$. This allows us to apply Lemma 19. Since there exists a $D_H \in \mathcal{D}_H$ consisting only of "high words", the inequalities are strict for at least one pair of D_H, D_L, where D_H is mapped to D_L under the randomized mapping. Using Lemma 19 and Lemma 9, an identical version of Lemma 10 exists and establishes stochastic dominance in this new model of preferences. □

The following remark establishes that the $(s_1^{\uparrow}, s_2^{\uparrow})$ strategy profile leads to a match on a "rarer" word than the $(s_1^{\downarrow}, s_2^{\downarrow})$ strategy profile. Specifically, we observe that for any pair of dictionaries D_1 and D_2, the $(s_1^{\uparrow}, s_2^{\uparrow})$ strategy profile yields an outcome at least as good as the $(s_1^{\downarrow}, s_2^{\downarrow})$ strategy profile from the system designer's perspective. In particular, when D_1 and D_2 overlap with more than one word, the $(s_1^{\uparrow}, s_2^{\uparrow})$ strategy profile yields the outcome of matching on the lowest frequency word and the $(s_1^{\downarrow}, s_2^{\downarrow})$ strategy profile yields the outcome of matching on highest frequency word. When D_1 and D_2 overlap with only one word, or no words, then $(s_1^{\uparrow}, s_2^{\uparrow})$ and $(s_1^{\downarrow}, s_2^{\downarrow})$ lead to the same outcome. The proof is easy and omitted.

Remark 1. *For any pair of dictionaries $D_1 \in \mathcal{D}_{e_1}$, $D_2 \in \mathcal{D}_{e_2}$, the $(s_1^{\uparrow}, s_2^{\uparrow})$ strategy profile yields a match on word with frequency at least as low as the $(s_1^{\downarrow}, s_2^{\downarrow})$ strategy profile.*

The following proposition follows from the fact that if player 2 plays H, player 1 maximizes the probability of matching by playing L, yet maximizes the probability of matching on the "rarest" word by playing H.

Proposition 6. $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ *is not an ordinal Bayesian-Nash equilibrium of the complete ESP game for any distribution over U under rare-words preferences.*

Proof. We show that the inequality in Definition 18 does not hold for $k = n$ when $(e_2, s_2) = (H, s_2^{\uparrow})$ and $(e_1, s_1) = (H, s_1^{\uparrow})$. Note that $\sum_{D_2} \Pr(D_2) \cdot I(D \cap D_2 \neq \emptyset) = \sum_{D_2} \Pr(D_2) \cdot I(g_w(s_1^{\uparrow}(D), s_2^{\uparrow}(D_2)) \in \{w_n, ..., w_1\})$. Thus showing that the inequality in Definition 18 does not hold for $k = n$ when $(e_2, s_2) = (H, s_2^{\uparrow})$ and $(e_1, s_1) = (H, s_1^{\uparrow})$ is equivalent to showing that Definition 15 does not hold for $k = d$ when $(e_2, s_2) = (H, s_2^{\downarrow})$ and $(e_1, s_1) = (H, s_1^{\downarrow})$, which is established by Lemma 10. $\square$

The implication of this proposition is that for every distribution over U, there exists a utility function for which $(H, s_1^{\uparrow})$ is not a best response to $(H, s_2^{\uparrow})$. Since $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ cannot be an ordinal Bayesian-Nash equilibrium for any distribution over U, we seek to understand under what conditions on utility functions and distributions we can get $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ as a Bayesian-Nash equilibrium. We restrict attention to the Zipfian distribution over U and multiplicative and additive valuation functions.

Definition 19. Let α_i denote the ratio of successive outcome in the utility function $v(o)$ (satisfying rare-words preferences) with total ordering of outcomes $o_1 \succ o_2 \succ ... \succ o_m$. That is, let $\alpha_i = \frac{v(o_i)}{v(o_{i+1})}$.

In order to prove positive results for the top level of the game under this new preference model, we use similar techniques as the previous section. In particular, we use a "randomized mapping", except in this case, we think of mapping each dictionary in $\mathcal{D}_L$ to a subset of dictionaries in $\mathcal{D}_H$. Rather than providing an intuitive explanation for the mapping as we did in the previous section, we show that a mapping with the desired properties exists.

Consider a mapping that maps each $D_L \in \mathcal{D}_L$ so some subset of dictionaries in $\mathcal{D}_H$. Suppose the dictionaries in $\mathcal{D}_L$ are indexed via $i = 1, ..., |\mathcal{D}_L|$ and the dictionaries in $\mathcal{D}_H$ are indexed via $j = 1, ..., |\mathcal{D}_H|$. Let the variable x_{ij} denote the probability that $D_{L,i}$ is mapped to $D_{H,j}$. Note that $x_{ij} = 0$ if and only if $D_{H,j} \cap U_L \nsubseteq D_{L,i}$. Therefore, we know that each $D_{L,i}$ is mapped to $\binom{|U_H| - |U_L| + d}{d}$. A valid mapping must satisfy the following properties:

1. $\sum_{j=1}^{|\mathcal{D}_H|} x_{ij} = 1 \; \forall i$ such that $1 \leq i \leq |\mathcal{D}_L|$
2. $\sum_{i=1}^{|\mathcal{D}_L|} \Pr(D_{L,i}) \cdot x_{ij} = \Pr(D_{H,j}) \; \forall j$ such that $1 \leq j \leq |\mathcal{D}_H|$

In order for the mapping to valid, it must also be the case that $0 \leq x_{ij} \leq 1$ for all i, j. Note that in the above sets of the equations, some of the x_{ij} are removed so that they will always be set to 0. A variable x_{ij} is removed from the above set of equations if and only if $D_{H,j} \cap U_L \nsubseteq D_{L,i}$.

In order to construct a solution to the above system of linear equations, we interpret the "randomized mapping" from the previous section as a system of linear equations.

1. $\sum_{i=1}^{|\mathcal{D}_L|} y_{ij} = 1\ \forall j$ such that $1 \leq j \leq |\mathcal{D}_H|$
2. $\sum_{j=1}^{|\mathcal{D}_H|} \Pr(D_{H,j}) \cdot y_{ij} = \Pr(D_{L,i})\ \forall i$ such that $1 \leq i \leq |\mathcal{D}_L|$

Similar to the previous system of linear equations, it must also be the case that $0 \leq y_{ij} \leq 1$ for all i, j, and some of the y_{ij} are removed so that they will always be set to 0. A variable y_{ij} is removed from the above set of equations if and only if $D_{H,j} \cap U_L \nsubseteq D_{L,i}$. The following two lemmas are proved in the Appendix and show that a solution exists to the first set of equations. Therefore, we know that a valid mapping exists from $\mathcal{D}_H$ to $\mathcal{D}_L$ with the desired properties.

Lemma 11. *There exists a solution to the second system of linear equations.*

Lemma 12. *If a solution to the second system of linear equations exists, then a solution to the first system of linear equations exists. Namely this solution can be obtained by* $x_{ij} = y_{ij} \cdot \frac{\Pr(D_{H,j}|H)}{\Pr(D_{L,i}|L)}$.

Lemma 13 states that the distribution obtained from sampling U_H directly is the same as the distribution obtained from sampling a low dictionary, followed by the randomized mapping. This follows immediately from the second set of conditions we require the randomized mapping to satisfy.

Lemma 13. $\Pr(D_{1,H}|H) = \sum_{D_{1,L}} \Pr(D_{1,L}|L) \cdot \Pr(g(D_{1,L}) = D_{1,H})$

We say that dictionary D' with elements $\{w'_d, w'_{d-1}, ..., w'_1\}$ (in order of increasing frequency) *dominates* dictionary D with elements $\{w_d, w_{d-1}, ..., w_1\}$ (in order of increasing frequency) if $f(w'_i) \leq f(w_i)$ for all i. We say that the dominance is *strict* if $D' \neq D$.

For the following lemmas, we use this randomized mapping, which satisfies the following property: Each $D_L \in \mathcal{D}_L$ is mapped to all dictionaries in $D_H \in \mathcal{D}_H$ such that $D_H \cap U_L \subseteq D_L$. In other words, each $D_L \in \mathcal{D}_L$ is mapped to dictionary $D_H \in \mathcal{D}_H$ with non-zero probability if and only if $D_H \cap U_L \subseteq D_L$. We show that if the randomized mapping satisfies this property, it must be the case that each *low dictionary* is mapped to a *high dictionary* that dominates it.

Lemma 14. *If* $D_H \subset \mathcal{D}_H$ *satisfies the property* $D_H \cap U_L \subseteq D_L$ *for any* $D_L \in \mathcal{D}_L$*, then each* D_H *dominates* D_L*. The dominance is strict when* $D_H \neq D_L$*.*

Proof. Let $D_L = \{w_d, w_{d-1}, ..., w_1\}$ and let $D_H = \{w'_d, w'_{d-1}, ..., w'_1\}$ (sorted in order of increasing frequency), where $D_H \cap U_L \subseteq D_L$. Assume D_H does not dominate D_L, so there exists a coordinate i such that $f(w_i) < f(w'_i)$. Let i be the minimum such coordinate. Since $D_L \subset U_L$, $w'_i \in U_L$, and so $w'_i \in D_L$. Since the dictionaries are in sorted order, $w_j = w'_i$ for some $j > i$, however this means there exists a $w'_k \in D_H \cap U_L$, where $w'_k \notin D_L$. Thus for all $1 \leq i \leq d$, $f(w_i) \geq f(w'_i)$. If $D_H \neq D_L$, there exists an i such that $w_i \neq w'_i$. Therefore, $f(w_i) > f(w'_i)$ and the dominance is strict. □

Lemma 15. *For every* $D_{1,L}$*, where* $D_{1,L}$ *is a dictionary sampled with respect to the* L *effort level, and for every* g *that satisfies the property that* $D_{1,L}$ *is mapped to a dictionary in* $D_{1,H} \in \mathcal{D}_H$ *such that* $D_{1,H} \cap U_L \subseteq D_{1,L}$ *and when both players play increasing frequency in the second stage and for* *all utility functions that satisfy rare-words preferences and* $\alpha_k \geq \frac{\Pr(w_{n-k} \in D_H)}{\Pr(w_{n-k+1} \in D_H)}$ *for all* k*, we have that:*

$$\sum_{D_{2,H}} \Pr(D_{2,H}) \cdot u(s_1^{\uparrow}(g(D_{1,L})), s_2^{\uparrow}(D_{2,H})) \geq \sum_{D_{2,H}} \Pr(D_{2,H}) \cdot u(s_1^{\uparrow}(D_{1,L}), s_2^{\uparrow}(D_{2,H})) \ \forall D_{1,L} \tag{9}$$

In addition, the inequality is strict when $g(D_{1,L}) \neq D_{1,L}$*.*

Lemma 16 uses Lemma 15 and Corollary 13 to show that playing H yields greater utility than L, given that the other playing is playing H, assuming players play increasing 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 16. *Given that players are playing words in order of increasing frequency,*

$$\sum_{D_{1,H}} \Pr(D_{1,H}|H) \sum_{D_{2,H}} \Pr(D_{2,H}|H) \cdot u(s_1^{\uparrow}(D_{1,H}), s_2^{\uparrow}(D_{2,H})) > \sum_{D_{1,L}} \Pr(D_{1,L}|L) \sum_{D_{2,H}} \Pr(D_{2,H}|H) \cdot u(s_1^{\uparrow}(D_{1,H}), s_2^{\uparrow}(D_{2,H})) \tag{10}$$

for all u *that satisfy rare-words preferences and* $\alpha_k \geq \frac{\Pr(w_{n-k} \in D_H)}{\Pr(w_{n-k+1} \in D_H)}$ *for all* k*.*

Theorem 4 together with Lemma 16 gives us the following result.

Proposition 7. $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ *is a Bayesian-Nash equilibrium of the ESP game for all distributions over* U *and any utility function that satisfies rare-words preferences and* $\alpha_k \geq \frac{\Pr(w_{n-k} \in D_H)}{\Pr(w_{n-k+1} \in D_H)}$ *for all* k*.*

Corollary 2. *For every distribution over* U*, there exists a utility function satisfying rare-words preferences for which* $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ *is a Bayesian-Nash equilibrium of the ESP game.*

Proposition 8. *There exists a distribution over* U *for which* $((H, s_1), (H, s_2))$ *cannot be a Bayesian-Nash equilibrium of the ESP game for any pair of consistent second-stage strategies* s_1, s_2 *and for any utility function satisfying match-early preferences.*

We interpret the conditions on the utility function in Theorem 7 for a specific class of distributions, namely the Zipfian distribution (see Definition 14). For this analysis, we restrict attention to the case where $s \leq 1$. Proposition 5 gives the criteria for $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ to be a Bayesian-Nash equilibrium for a family of Zipfian distributions, when the dictionaries are sampled with replacement. We note that for large values of n the conditions for when the dictionaries are sampled without replacement are virtually identical to the case where the dictionaries are sampled with replacement.

Definition 20. Additive Discount Property: Under rare-words preferences, a utility function $v(o)$ over the total ordering of outcomes $o_1 \succ o_2 \succ ... \succ o_m$ satisfies the additive discount property if and only if, for each pair of adjacent outcomes o_j and o_{j+1}, $v(o_j) - v(o_{j+1}) = c$ for some constant $c > 0$ and $v(o_m) = 0$.

Definition 21. Multiplicative Discount Property: Under rare-words preferences, a utility function $v(o)$ over the total ordering of outcomes $o_1 \succ o_2 \succ ... \succ o_m$ satisfies the multiplicative discount property if and only if, for each pair of adjacent outcomes o_j and o_{j+1}, $\frac{v(o_j)}{v(o_{j+1})} \geq r$ for some constant $r > 1$.

Theorem 5. *$((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ is a Bayesian-Nash equilibrium of the complete ESP game for Zipfian distributions over U with $s \leq 1$ and any additive utility function that satisfies rare-words preferences and any multiplicative utility function that satisfies rare-words preferences with $r \geq 2$.*

Proof. From Lemma 15, it suffices to show $t_i' \cdot v(w_i') > t_i \cdot v(w_i)$. First consider the case of additive utility functions. Let us assume that w_i' is the j^{th} most frequent word in the universe and w_i is the k^{th} most frequent word in the universe (with $j > k$). Since words are sampled according to the Zipfian distribution, $t_i' = 1 - (1 - \frac{1}{j^s Z})^d$ and $t_i = 1 - (1 - \frac{1}{k^s Z})^d$, where $Z = \sum_{w \in U_M} f(w) - \sum_{j=d}^{i+1} f(w_j)$. We know according to Lemma 18 that $\frac{t_i}{t_i'} \leq \frac{j^s}{k^s} \leq \frac{j}{k}$. Since $\frac{t_i}{t_i'} \leq \frac{j}{k}$, we have $\frac{t_i}{t_i'} \leq \frac{j \cdot c}{k \cdot c}$ which gives us $t_i' \cdot v(w_i') > t_i \cdot v(w_i)$ for all additive utility functions.
Now we consider the case of multiplicative utility functions. We have shown that $\frac{t_i}{t_i'} \leq \frac{j}{k}$. Since $\frac{j}{k} \leq \frac{2^j}{2^k}$ for all integers $j > k$, $t_i' \cdot v(w_i') > t_i \cdot v(w_i)$ for all multiplicative utility functions with $r \geq 2$. □

5. Conclusions

In this paper, we introduced a simple model of the ESP game and provided, in many cases, complete equilibrium characterizations. We introduced the model of match-early preferences to capture the current implementation of the ESP game. We showed that the strategy profile (s, s), where s is an *almost decreasing* strategy, is a strict ordinal Bayesian-Nash equilibrium of the second-stage of the ESP game (under match-early preferences) for every distribution over U, irrespective of the effort levels chosen in the first-stage. Moreover, we showed that these are the only strategy profiles, where at least one player is playing a consistent strategy, that is a strict ordinal Bayesian-Nash equilibrium for every distribution over U. It is interesting to note that these results hold even if players have different distributions over U, as long as the total ordering of words in U in terms of frequency is the same for both players. Since the $(s_1^{\downarrow}, s_2^{\downarrow})$ is the most natural of the set of strategies that satisfy the almost decreasing property, we focused on this equilibrium profile when analyzing the equilibrium of the complete game.

The implication of equilibrium characterization for the second-stage ESP game under match-early preferences is that there exists a distribution such that $(s_1^{\uparrow}, s_2^{\uparrow})$ cannot be an ordinal Bayesian-Nash equilibrium. However, we can make a stronger claim, that there exist distributions for which $(s_1^{\uparrow}, s_2^{\uparrow})$ cannot be a Bayesian-Nash equilibrium for any valuation function satisfying match-early preferences. We showed that the Zipfian distribution is one such distribution. The Zipfian distribution is significant in this setting since the distribution of words in the English language is known to follow a Zipfian with exponent very close to 1 [31].

Given the equilibrium analysis for the second-stage, we showed that $((L, s_1^{\downarrow}), (L, s_2^{\downarrow}))$ is a strict ordinal Bayesian-Nash equilibrium of the complete ESP game, under match-early preferences, for every distribution over U, except for the uniform distribution. We precluded the existence of a $((H, s_1^{\downarrow}), (H, s_2^{\downarrow}))$ equilibrium for any distribution over U, except for the uniform distribution, and any utility function that satisfies match-early preferences, by showing that $(L, s_1^{\downarrow})$ is a strict ordinal best response to $(H, s_2^{\downarrow})$. In the case of the uniform distribution over U, we established both

$((L, s_1^{\downarrow}), (L, s_2^{\downarrow}))$ and $((H, s_1^{\downarrow}), (H, s_2^{\downarrow}))$ are weak ordinal Bayesian-Nash equilibrium of the complete ESP game. While the model of the ESP game states that users have the same valuation, we note that in establishing ordinal Bayesian-Nash equilibrium, these strategy profiles are in equilibrium for any pair of player valuations that satisfy match-early preferences. In other words, the two player valuations need not be the same for the strategy profile $((L, s_1^{\downarrow}), (L, s_2^{\downarrow}))$ to be a Bayesian-Nash equilibrium. Our equilibrium analysis supports the empirical results and observations of others that users tend to coordinate on low effort words under match-early preferences.

In order to remedy the problem of users coordinating on easy words, we introduced the rare-words preferences model. We showed that $(s_1^{\uparrow}, s_2^{\uparrow})$ is a strict ex-post Nash equilibrium for the second-stage of the ESP game under rare-words preferences and every distribution over U. Therefore, we focused on the $(s_1^{\uparrow}, s_2^{\uparrow})$ strategy profile when analyzing the complete game. We also showed that the strategy $s^{\downarrow}$ is strictly dominated for the second-stage of the ESP game under rare-words preferences and every distribution over U. This is in contrast with the equilibrium analysis of the second-stage game under match-early preferences, where we showed that $(s_1^{\downarrow}, s_2^{\downarrow})$ is one of only a few strategy profiles that are ordinal Bayesian-Nash equilibrium for every distribution over U. We also showed that $(s_1^{\uparrow}, s_2^{\uparrow})$ is not an ordinal Bayesian Nash equilibrium for every distribution over U and there exist distributions, including the Zipfian distribution, such that $(s_1^{\uparrow}, s_2^{\uparrow})$ cannot be a Bayesian-Nash equilibrium for any valuation function consistent with match-early preferences.

We analyzed the complete ESP game under rare-words preferences. We found that $((L, s_1^{\uparrow}), (L, s_2^{\uparrow}))$ is a strict ordinal Bayesian-Nash equilibrium for ever distribution over U. While this may not be ideal, we found that the strategy profile $((L, s_1^{\uparrow}), (L, s_2^{\uparrow}))$ leads to a (weakly) better outcome than the strategy profile $((L, s_1^{\downarrow}), (L, s_2^{\downarrow}))$, which was the equilibrium under match-early preferences. We also found that $((H, s_1^{\uparrow}), (H, s_2^{\uparrow})$ is not an ordinal Bayesian-Nash equilibrium for any distribution over U. On the positive side, we showed that $((H, s_1^{\uparrow}), (H, s_2^{\uparrow}))$ can be a Bayesian-Nash equilibrium for every distribution over U, for certain conditions on the utility function. Since the distribution of words in the English language is known to follow a Zipfian distribution with exponent very close to 1, we interpreted what these conditions mean for the Zipfian distribution by focusing attention on two very natural classes of utility functions, additive utility functions and multiplicative utility functions.

Proposition 5 showed an interesting result in that, in order to have coordination on low frequency, more descriptive words, it suffices to have a constant difference in points between each subsequent word in the universe. A key challenge in implementing this result from the system designer's perspective is that she does not know the set of words in the universe *a priori*. A natural way around this problem is for the center to update scores as it learns more information about the set of words in the universe. That is, the center can award points based on the current knowledge of the universe and update the point total as the set of words in the universe grows through subsequent game play. Though it is perhaps unsatisfying to not be able to award the full amount of points upon agreement, it is important to note that the point total will only increase as the center learns more about the set of words in the universe. Another way to implement this point scheme would be to award points with the set of words in the English language as the point total. However, this could potentially lead to very large point differences between matching on successive words in the universe. The advantage of this scheme is that there would be no need for updating point totals after the fact.

References

[1] Adamic, L.A., Zhang, J., Bakshy, E., Ackerman, M.S., 2008. Everyone knows something: Examining knowledge sharing on Yahoo answers, in: Proc. 17th Intl. Conf. on the World Wide Web (WWW).

[2] von Ahn, L., 2006. Games with a purpose. IEEE Computer 39, 92–94.

[3] von Ahn, L., Dabbish, L., 2004. Labeling images with a computer game, in: Proceedings of the 2004 Conference on Human Factors in Computing Systems (CHI), pp. 319–326.

[4] von Ahn, L., Dabbish, L., 2008. Designing games with a purpose. Communications of the ACM 51, 58–67.

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

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

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

[8] Archak, N., Sundararajan, A., 2009. Optimal design of crowdsourcing contests, in: ICIS.

[9] Barnard, K., Duygulu, P., Forsyth, D., de Freitas, N., Blei, D.M., Jordan, M.I., 2003. Matching words and pictures. Journal of Machine Learning Research 3, 1107–1135.

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

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

[12] Chawla, S., Hartline, J., Sivan, B., 2011. Optimal crowdsourcing contests, in: Workshop on Social Computing and User Generated Contest.

[13] d'Aspremont, C., Peleg, B., 1988. Ordinal bayesian incentive compatible representation of committees. Social Choice and Welfare 5, 261–280.

[14] DiPalatino, D., Vojnovic, M., 2009. Crowdsourcing and all-pay auctions, in: ACM Conference on Electronic Commerce.

[15] Ghosh, A., Hummel, P., 2011. A game-theoretic analysis of rank order mechanisms for user generated content, in: ACM Conference on Electronic Commerce (EC).

[16] Ghosh, A., McAfee, P., 2011. Incentivizing high quality user generated content, in: WWW.

[17] Hacker, S., von Ahn, L., 2009. Matchin: Eliciting user preferences with an online game, in: Proc. SIGCHI Conf. on Human Factors in Computing Systems.

[18] Hanoch, G., Levy, H., 1969. The efficiency analysis of choices involving risk. Review of Economic Studies 36, 335–346.

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

[20] Ho, C.J., Chang, T.H., Lee, J.C., jen Hsu, J.Y., Chen, K.T., 2009. Kisskissban: A competitive human computation game for image annotation, in: KDD-HCOMP.

[21] Ho, C.J., Chen, K.T., 2009. On formal models for social verification, in: KDD-HCOMP.

[22] Jain, S., Chen, Y., Parkes, D.C., 2009. Designing incentives for online question and answer forums, in: ACM Conference on Electronic Commerce.

[23] Kearns, M., Suri, S., Montfort, N., 2006. An experimental study of the human coloring problem on human subject networks. Science 313, 824–827.

[24] Law, E., von Ahn, L., 2009. Input-agreement: A new mechanism for data collection using human computation games, in: Proc. SIGCHI Conf. on Human Factors in Computing Systems.

[25] Majumdar, D., 2003. Ordinally bayesian incentive compatible stable matching. Working Paper.

[26] Moldovanu, B., Sela, A., 2001. The optimal allocation of prizes in contests. American Economic Review 91, 542–558.

[27] Moldovanu, B., Sela, A., 2006. Contest architecture. Journal of Economic Theory 126, 70–96.

[28] Nam, K.K., Ackerman, M.S., Adamic, L.A., 2009. Questions in, Knowledge iN?: A study of Naver's question answering community, in: Proc. SIGCHI Conf. on Human Factors in Computing Systems (CHI).

[29] Weber, I., Robertson, S., Vojnovic, M., 2008. Rethinking the ESP Game. Technical Report. Microsoft Research.

[30] Yang, J., Adamic, L.A., Ackerman, M.S., 2008. Crowdsourcing and knowledge sharing: Strategic user behavior on Taskcn, in: Proc. 9th ACM Conf. on Electronic Commerce (EC).

[31] Zipf, G., 1932. Selective Studies and the Principle of Relative Frequency in Language. Harvard Univ. Press.

Appendix A. Appendix

Lemma 17. *If dictionary D and dictionary D' only differ by one element, x_i and x'_i respectively, with $f_{e_1}(x_i) < f_{e_2}(x'_i)$, then dictionary D' is sampled with strictly greater probability than dictionary D as long as $e_1 \geq e_2$.*

Proof. If $e_1 > e_2$, $\Pr(D'|e_2) > \Pr(D'|e_1)$. Therefore it suffices to show this for $e_1 = e_2$. A particular dictionary can be sampled in one of $d!$ ways. Each permutation of D has a corresponding permutation of D' that involves replacing x_i with x'_i. Let $A = a_1, ..., a_d$ be a permutation of D and let $A' = a'_1, ..., a'_d$ be the corresponding permutation of D', where A and A' differ in coordinate j. A is sampled with probability $\Pr(a_1)\Pr(a_2|a_1)...\Pr(a_d|a_1, a_2, ..., a_{d-1})$ and A' is sampled with probability $\Pr(a'_1)\Pr(a'_2|a'_1)...\Pr(a'_d|a'_1, a'_2, ..., a'_{d-1})$. We know $\Pr(a_k|a_1, ..., a_{k-1}) = \Pr(a'_k|a'_1, ..., a'_{k-1})$ for all $k < j$ and $\Pr(a_k|a_1, ..., a_{k-1}) < \Pr(a'_k|a'_1, ..., a'_{k-1})$ for all $k \geq j$. Hence, for each permutation of D, there exists a corresponding permutation of D' that is sampled with strictly greater probability and $\Pr(D'|e_1) > \Pr(D|e_1)$. □

Lemma 3. If a consistent strategy s_2 satisfies the almost decreasing property, then the strategy profile (s_2, s_2) is a strict ordinal Bayesian-Nash equilibrium of the second-stage ESP game under match-early preferences, for every choice of effort levels e_1 and e_2, for every distribution over U.

Proof. For each w'_i, w'_j such that $w'_i \succ w'_j$ under s_2, where s_2 is an almost decreasing strategy, $f(w'_i) \geq f(w'_j)$, as long as w'_j is not the least priority word under s_2, for all distributions over U. Consider the set A 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 B 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 $\max(1, i-n+d) \leq k \leq \max(i, d-1)$. If w'_j is the least priority word under s, $B = \emptyset$ and $A \neq \emptyset$, so $\Pr(B) < \Pr(A)$. Now suppose w'_j is not the least priority element under s_2. Notice B is the exactly the set of dictionaries that satisfy: $w'_j \in l_{\leq k}(s_2(D_2)) \cap w'_i \notin D_2$. There exists a mapping $t : B \rightarrow A$, which takes a $b \in B$ to an $a \in A$, by removing w'_j and replacing it with w'_i. The mapping t takes each element $b \in B$ to a unique element in $a \in A$, where $\Pr(b) < \Pr(a)$, due to Lemma 17. Therefore $Pr(B) < Pr(A)$. Hence the *preservation condition* is satisfied when s_2 is an almost decreasing strategy, regardless of distribution. Finally, when s_2 is an almost decreasing strategy, we have that $\Pr(w'_i \in D_2) > \Pr(w'_j \in D_2)$, for all j, i such that $n > j > i$. Likewise, $\Pr(w'_j \in D_2) \geq \Pr(w'_j \in l_{\leq k}(s_2(D_2))$, for all j, k. This gives us: $\Pr(w'_i \in D_2) > Pr(w'_j \in D_2) \geq \Pr(w'_j \in l_{\leq k}(s_2(D_2)))$ for all i, j, k where $i+1 < j < n$ and $k \leq d-1$ and for all distributions. When $j = n$, $\Pr(w'_n \in l_{\leq k}(s_2(D_2)) = 0$ for all $k \leq d-1$, so $\Pr(w'_i \in D_2) > \Pr(w'_n \in l_{\leq k}(s_2(D_2))$ for all $i < n$ and $k \leq d-1$. Hence, the *strong condition* is satisfied for all distributions over U. Lemma 2, along with Theorem 1, gives us the desired result. □

Lemma 5. For every (e_2, s_2) (except for s_2 that are almost decreasing), there exists a distribution over U, an effort level e_1, and a dictionary for which Algorithm 1 will not output an ordering in agreement with s_2.

Proof. Since s_2 is not an almost decreasing strategy, there exists adjacent $w'_i \succ w'_{i+1}$ under s_2 such that $f(w'_i) < f(w'_{i+1})$ and $i < n-1$. Let i be the smallest such index. Assume that w'_{i+1} is the k^{th} most frequent element in U. Consider the following distribution over U: The top k most frequent words in U have frequency $\frac{1-\epsilon}{k}$ and the $n-k$ least frequent words in U have frequency $\frac{\epsilon}{n-k}$, where $\epsilon < \frac{(n-2d+1)(n-k)}{nk-dk+(n-k)(n-2d+1)}$. (The RHS > 0 since $d \leq \frac{n}{2}$.) Consider the following sets, $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$: Set $\mathcal{A}$ contains dictionaries that have w'_i in the top j positions of $s_2(D_2)$ (where $\max(1, i-n+d) \leq j \leq \max(i, d-1)$) and do not contain w'_{i+1}. Set $\mathcal{B}$ contains dictionaries that have w'_i in the top j positions of $s_2(D_2)$ and also contain w'_{i+1}. Set $\mathcal{C}$ contains dictionaries that have have w'_{i+1} in the top i positions and do not contain w'_i. We construct $t_1 : \mathcal{A} \to \mathcal{C}$ and $t_2 : \mathcal{B} \to \mathcal{C}$. t_1 replaces w'_i with w'_{i+1}. Since w'_i and w'_{i+1} are adjacent in s_2 and $\mathcal{A}$ and $\mathcal{C}$ are non-empty, t_1 is a bijection. From Lemma 17, each $A \in \mathcal{A}$ that occurs with probability p_A is mapped to a $C \in \mathcal{C}$ that occurs with probability at least as high as $p_A \cdot \frac{f(w_{i+1})}{f(w_i)}$. There exists at least one case such that this inequality is strict. Thus $\Pr(D_2 \in \mathcal{C}) > \frac{f(w_{i+1})}{f(w_i)} \cdot \Pr(D_2 \in \mathcal{A})$. Note that $|\mathcal{B}| = \binom{n-2}{d-2}$ and $|\mathcal{C}| = \binom{n-2}{d-1}$. Since $d \leq \frac{n}{2}$, $\binom{n-2}{d-2} < \binom{n-2}{d-1}$, so $|\mathcal{B}| < |\mathcal{C}|$. Thus there exists a t_2 that takes each $B \in \mathcal{B}$ to a unique $C \in \mathcal{C}$. t_2 removes w'_i and replaces it with an $x \in U - B$. t_2 maps a $B \in \mathcal{B}$ that occurs with probability p_B to a $C \in \mathcal{C}$ that occurs with probability $p_C \geq p_B$. Therefore, $\frac{\Pr(D_2 \in \mathcal{B})}{\Pr(D_2 \in \mathcal{C})} \leq \frac{\binom{n-2}{d-2}}{\binom{n-2}{d-1}}$. Combining this with $\frac{\Pr(D_2 \in \mathcal{A})}{\Pr(D_2 \in \mathcal{C})} < \frac{f(w'_i)}{f(w'_{i+1}} < \frac{n-2d+1}{n-d}$, $\Pr(D_2 \in \mathcal{A}) + \Pr(D_2 \in \mathcal{B}) < \Pr(D_2 \in \mathcal{C})$. This implies $\Pr(w'_i \in l_{\leq j}(s_2(D_2))) < \Pr(w'_{i+1} \in l_{\leq j}(s_2(D_2)))$. Consider any D_1 with w'_i as the j^{th} highest priority word under s_2. Given the selection of i, the first $j-1$ words of D_1 are in order of decreasing frequency. Therefore, Algorithm 1 will output the first $j-1$ words of D_1 according to s_2. At the j^{th} step of the algorithm, since $\Pr(w'_i \in l_{\leq j}(s_2(D_2))) < \Pr(w'_{i+1} \in l_{\leq j}(s_2(D_2)))$, w'_i will not be output. □

Lemma 6. If there exists an $1 \leq i \leq n$ such that $\Pr(w'_i \notin l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_{i+1} \in l_{\leq j}(s_2^{\uparrow}(D_2))) \geq \Pr(w'_i \in l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_{i+1} \notin l_{\leq j}(s_2^{\uparrow}(D_2)))$ for some $\max(1, i-n+d) \leq j \leq \max(i, d-1)$, then $(\uparrow, \uparrow)$ cannot be a Bayesian-Nash equilibrium for any valuation function satisfying match-early preferences.

Proof. Suppose $s_1 = s_1^{\uparrow}$ and $s_2 = s_2^{\uparrow}$. Consider adjacent w'_i and w'_{i+1} with $f(w'_i) > f(w'_{i+1})$ and consider a D_1 that contains w'_i and w'_{i+1}. We consider player 1's utility of playing $s_1^{\uparrow}(D_1)$ versus the utility of playing the same ordering yet swapping w'_i and w'_{i+1}. Call the latter strategy s'_1. Assume that w'_i is the j^{th} word of $s_1^{\uparrow}(D_1)$. Consider the following cases:

1. Suppose the match happens before l_j or after l_{j+1}, the swap does not change the outcome.
2. Suppose the match happens in l_j. If the match happens does not happen on w'_i, the swap does not change the outcome. If the match happens on w'_i and $w'_{i+1} \notin D_2(j)$, the swap leads to payoffs of $v(l_{j+1})$ instead of $v(l_j)$. Otherwise, if $w'_{i+1} \in D_2(j)$, the swap does not change the outcome.
3. Suppose the match happens in l_{j+1}. If the match does not happen on w'_{i+1}, the swap does not change the outcome. If the match happens on w'_{i+1} and $w'_{i+1} \in D_2(j)$, the swap leads to payoffs of $v(l_{j+1})$ instead of $v(l_j)$. Otherwise, if w'_{i+1} is the $j+1^{st}$ word of player 2, the swap does not change the outcome.

Therefore, $u(s_1^{\uparrow}(D_1), s_2^{\uparrow}) - u(s'_1(D_1), s_2^{\uparrow}) = \Pr(w'_i \in l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_{i+1} \notin l_{\leq j}(s_2^{\uparrow}(D_2))) \cdot (v(l_j) - v(l_{j+1})) - \Pr(w'_{i+1} \in l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_i \notin l_{\leq j}(s_2^{\uparrow}(D_2))) \cdot (v(l_j) - v(l_{j+1}))$. This expression will be negative for some $D_1 \in \mathcal{D}_1$ if and only if there exists a value of $\max(1, i-n+d) \leq j \leq \max(i, d-1)$ such that $\Pr(w'_i \in l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_{i+1} \notin l_{\leq j}(s_2^{\uparrow}(D_2))) < \Pr(w'_{i+1} \in l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_i \notin l_{\leq j}(s_2^{\uparrow}(D_2)))$. □

Proposition 2. Second-stage strategy profile $(\uparrow, \uparrow)$ cannot be a Bayesian-Nash equilibrium for the second-stage of the ESP game for any Zipfian distribution over U with $s \leq 1$ and for any valuation function satisfying match-early preferences.

Proof. From Lemma 6, it suffices to show there exists $1 \leq i \leq d$ such that $\Pr(w'_i \notin l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_{i+1} \in l_{\leq j}(s_2^{\uparrow}(D_2))) > \Pr(w'_i \in l_{\leq j}(s_2^{\uparrow}(D_2)) \cap w'_{i+1} \notin l_{\leq j}(s_2^{\uparrow}(D_2)))$ for some $\max(1, i-n+d) \leq j \leq \max(i, d-1)$. Consider D_1 consisting of the d lowest frequency words for player 2. We show that $\Pr(w'_{d-1} \notin l_{\leq d-1}(s_2^{\uparrow}(D_2)) \cap w'_d \in l_{\leq d-1}(s_2^{\uparrow}(D_2))) \geq \Pr(w'_{d-1} \in l_{\leq d-1}(s_2^{\uparrow}(D_2)) \cap w'_d \notin l_{\leq d-1}(s_2^{\uparrow}(D_2)))$. This is equivalent to $\Pr(w'_{d-1} \in l_{\leq d-1}(s_2(D_2)) \cap w'_d \notin D_2) + \Pr(w'_{d-1} \in l_{d-1}(s_2(D_2)) \cap w'_d \in l_d(s_2(D_2))) < \Pr(w'_d \in l_{\leq d-1}(s_2(D_2)) \cap w'_{d-1} \notin D_2)$. Define $\mathcal{A}$ as the set of D_2 that satisfy: $w'_{d-1} \in l_{\leq d-1}(s_2(D_2)) \cap w'_d \notin D_2$. Define B as the set of D_2 that satisfy: $w'_{d-1} \in l_{d-1}(s_2(D_2)) \cap w'_d \in l_d(s_2(D_2))$ and finally, define $\mathcal{C}$ as the set of D_2 that satisfy: $w'_d \in l_{\leq d-1}(s_2(D_2)) \cap w'_{d-1} \notin D_2$. Construct $t_1 : \mathcal{A} \to \mathcal{C}$ that removes w'_{d-1} from $A \in \mathcal{A}$ and replacing it with w'_d in $C\mathcal{C}$. Since w'_{d-1} and w'_d are adjacent in s_2 and $\mathcal{A}$ and $\mathcal{C}$ are non-empty, t_1 is a bijection. Each $A \in \mathcal{A}$ that occurs with probability p_A is mapped to a $C \in \mathcal{C}$ that occurs with probability greater than $p_A \cdot \frac{f(w'_d)}{f(w'_{d-1})}$ (from Lemma 17). Construct t_2 that maps each $B \in \mathcal{B}$ to a $C \in \mathcal{C}$, by removing w'_{d-1} and replacing it with w'_n. Therefore, each $B \in \mathcal{B}$ that occurs with probability p_B is mapped to a $C \in \mathcal{C}$ that occurs with probability greater than $p_A \cdot \frac{f(w'_n)}{f(w'_{d-1})}$. Since w'_{d-1} and w'_d are the $n-d+2$ and $n-d+1$ most frequent elements in U, this gives us $\frac{\Pr(A)}{\Pr(C)} < \frac{f(w'_{d-1})}{f(w'_d)} = \frac{(n-d+1)^s}{(n-d+2)^s} \leq \frac{n-d+1}{n-d+2}$ and $\frac{\Pr(B)}{\Pr(C)} < \frac{f(w'_{d-1})}{f(w'_n)} = \frac{1}{(n-d+2)^s} \leq \frac{1}{n-d+2}$. Therefore, $\frac{\Pr(A)}{\Pr(C)} + \frac{\Pr(B)}{\Pr(C)} < 1$, which establishes the desired result. □

Theorem 4. Second-stage strategy profile $(s_1^{\uparrow}, s_2^{\uparrow})$ is a strict ex-post Nash equilibrium for the second-stage of the ESP game for every distribution over U and every $e_1 = e_2$, under rare-words preferences.

Proof. Given $s_2 = s_2^{\uparrow}$, we need to show that for any $s_1 \neq s_1^{\uparrow}$, player 1 (weakly) prefers $s_1^{\uparrow}$ to s_1 for all D_1, D_2 and player 1 strongly prefers $s_1^{\uparrow}$ to s_1 for some D_1, D_2. Since $s_1 \neq s_1^{\uparrow}$, there exists a D_1 for which the sampled ordering: $w'_1 \succ w'_2 \succ ... \succ w'_d$, contains a pair of adjacent $w'_i \succ w'_{i+1}$ with $f(w'_i) > f(w'_{i+1})$. We assume i is the smallest index that satisfies this property. Consider the following cases:
1. Suppose the match occurs before l_i or after l_{i+1}, swapping w'_i and w'_{i+1} does not change the outcome.
2. Suppose the match occurs at l_i. If the match happens on some $w \neq w'_i$, then $f(w) < f(w'_i)$. If $w'_{i+1} \in l_{\leq i}(s_2^{\uparrow}(D_2))$ and $f(w'_{i+1}) < f(w)$, then swapping w'_i and w'_{i+1} leads to a strictly better outcome. Otherwise, swapping does not change the outcome. If instead, the match occurs on w'_i and $w'_{i+1} \in l_{\leq i}(s_2^{\uparrow}(D_2))$, w'_{i+1} is the word matched upon, which is a strictly better outcome. If $w'_{i+1} \notin l_{\leq i}(s_2^{\uparrow})$, w'_i is matched upon in $l_i + 1$ and player 1 is indifferent between the two outcomes.
3. Suppose the match occurs at l_{i+1}. If the match happens on some $w \neq w'_{i+1}$, then $w \in l_{i+1}(s_2(D_2))$ and swapping w'_i and w'_{i+1} does not change the outcome. If the match happens on w'_{i+1}, swapping w'_i and w'_{i+1} yields an outcome of matching on w'_{i+1} in l_i, but player 1 is indifferent between these two outcomes.
If $s_2^{\uparrow}(D_2)$ has w'_{i+1} in position i, w'_i in position $i+1$ and $l_{\leq i}(s_1(D_1)) \cap l_{\leq i}(s_2^{\uparrow}(D_2)) = \emptyset$, then player 1 strongly prefers $s_1^{\uparrow}$ to any other s_1. □

Proposition 4. Second-stage strategy $s_1^{\downarrow}$ is strictly dominated for any second-stage strategy of player 2 and for any distribution over U and any choice of effort levels e_1, e_2, under rare-words preferences.

Proof. We will exhibit a strategy that (weakly) dominates $s_1^{\downarrow}$ for any D_1, D_2 and s_2 and strictly dominates $s_1^{\downarrow}$ for some D_1, D_2 and s_2. Suppose $s_1 = s_1^{\downarrow}$. Let w_1' and w_2' be the two highest frequency words of player 1, with $f(w_1') > f(w_2')$. Consider the following cases:
1. Suppose the match happens after l_2. Swapping the first two words does not change the outcome.
2. Suppose the match happens on w_1'. Now suppose the match occurs in l_1 and $w_2' = l_2(s_2(D_2))$, swapping w_1' and w_2' leads to a strictly better outcome: w_2' is matched upon instead of w_1'. If the match occurs in l_1 and $w_2' \neq l_2(s_2(D_2))$, swapping w_1' and w_2' does not change the outcome. If the match happens in l_2, swapping w_1' and w_2' does not change the outcome.
3. Suppose the match happens on w_2' in l_2. If $w_2' \in l_1(s_2(D_2))$, then swapping improves the outcome, otherwise it does not. □

Lemma 11. There exists a solution to the second system of linear equations.

Proof. We define the candidate solution to the linear system of equations as follows: For all i, j, if $D_{H,j} \cap U_L \nsubseteq D_{L,i}$, $y_{ij} = 0$.
Otherwise, (that is, $D_{H,j} \cap U_L \subseteq D_{L,i}$),

$$y_{ij} = \frac{\Pr_{U_L^d}(D_{L,i})}{\sum_{D_{L,k}:D_{H,j}\cap U_L \subseteq D_{L,k}} \Pr_{U_L^d}(D_{L,k})}$$

where $\Pr_{U_L^d}(D_{L,i}) = \Pr(D_{L,i})$ denotes the probability of obtaining $D_{L,i}$, when sampling from the universe U_L d times without replacement. (Most of the time the U_L^d notation is implied, but for the purposes of this proof it is necessary for clarity).
It is easy to see from this definition of y_{ij} that $0 \leq y_{ij} \leq 1$ for all i, j. Also note that, for all j:

$$\sum_{i=1}^{|\mathcal{D}_L|} y_{ij} = \sum_{D_{L,k}:D_{H,j}\cap U_L \subseteq D_{L,k}} \frac{\Pr(D_{L,k})}{\sum_{D_{L,k'}:D_{H,j}\cap U_L \subseteq D_{L,k'}} \Pr(D_{L,k})} = 1$$

Now it remains to show that the candidate y_{ij} satisfies equations of the second type.
Note that:

$$\begin{aligned}
&\frac{\Pr_{U_L^d}(D_{L,i})}{\sum_{D_{L,k}:D_{H,j}\cap U_L \subseteq D_{L,k}} \Pr_{U_L^d}(D_{L,k})} \\
&= \frac{\Pr(D_{H,j}\cap U_L)\Pr(D_{L,i}\setminus D_{H,j}|D_{H,j}\cap U_L)}{\sum_{D_{L,k}:D_{H,j}\cap U_L \subseteq D_{L,k}} \Pr(D_{H,j}\cap U_L)\Pr(D_{L,k}\setminus D_{H,j}|D_{H,j}\cap U_L)} \\
&= \frac{\Pr(D_{L,i}\setminus D_{H,j}|D_{H,j}\cap U_L)}{\sum_{D_{L,k}:D_{H,j}\cap U_L \subseteq D_{L,k}} \Pr(D_{L,k}\setminus D_{H,j}|D_{H,j}\cap U_L)} \\
&= \frac{\Pr_{(U_L\setminus D_{H,j})^{d-a}}(D_{L,i}\setminus D_{H,j})}{\sum_{D_{L,k}\setminus D_{H,j}:D_{H,j}\cap U_L \subseteq D_{L,k}} \Pr_{(U_L\setminus D_{H,j})^{d-a}}(D_{L,k}\setminus D_{H,j})} \qquad \text{(A.1)}
\end{aligned}$$

where $a = |D_{H,j} \cap U_L|$ (e.g. the number of "low words" in $D_{H,j}$) and $\Pr_{(U_L\setminus D_{H,j}))^{d-a}}(S)$ denotes the probability of obtaining the set S from $d-a$ samples from the set $U_L \setminus D_{H,j}$, where the sampling is done without replacement. It is easy to see that the denominator of Eq. 7 equals 1. This gives us:

$$y_{ij} = \Pr_{(U_L\setminus D_{H,j})^{d-a}}(D_{L,i}\setminus D_{H,j}) = \Pr_{(U_H\setminus D_{H,j})^{d-a}}(D_{L,i}\setminus D_{H,j})$$

where the last expression denotes the probability that you get $D_{L,i} \setminus D_{H,j}$ when sampling from $U_H \setminus D_{H,j}$ until you get $d-a$ low words (e.g. $d-a$ words from U_L). Now going back to equations of the second type:
$\sum_{j=1}^{|\mathcal{D}_H|} \Pr(D_{H,j}) \cdot y_{ij} = \sum_{j=1}^{|\mathcal{D}_H|} \Pr_{U_H^d}(D_{H,j}) \cdot \Pr_{(U_H \setminus D_{H,j})^{d-a}}(D_{L,i} \setminus D_{H,j}) = \Pr_{U_L^d}(D_{L,i})$ for all i. □

Lemma 12. If a solution to the second system of linear equations exists, then a solution to the first system of linear equations exists. Namely this solution can be obtained by $x_{ij} = y_{ij} \cdot \frac{\Pr(D_{H,j}|H)}{\Pr(D_{L,i}|L)}$.

Proof. Since a solution exists to the second system of linear equations, we know that there exists a set of y_{ij} that satisfies: $\sum_{j=1}^{|\mathcal{D}_H|} \Pr(D_{H,j}) \cdot y_{ij} = \Pr(D_{L,i})\ \forall i$ such that $1 \leq i \leq |\mathcal{D}_L|$, or in other words, $\sum_{j=1}^{|\mathcal{D}_H|} \frac{\Pr(D_{H,j})}{\Pr(D_{L,i})} \cdot y_{ij} = 1\ \forall i$ such that $1 \leq i \leq |\mathcal{D}_L|$. Observe that this gives a solution to the first type of equation in the first linear system of equations, where $x_{ij} = y_{ij} \cdot \frac{\Pr(D_{H,j})}{\Pr(D_{L,i})}$.
We also know that this same set of y_{ij} satisfies: $\sum_{i=1}^{|\mathcal{D}_L|} y_{ij} = 1\ \forall j$ such that $1 \leq j \leq |\mathcal{D}_H|$. If $y_{ij} = x_{ij} \cdot \frac{\Pr(D_{L,i})}{\Pr(D_{H,j})}$, then we know that $\sum_{i=1}^{|\mathcal{D}_L|} x_{ij} \cdot \frac{\Pr(D_{L,i})}{\Pr(D_{H,j})} = 1\ \forall j$ such that $1 \leq j \leq |\mathcal{D}_H|$, or in other words, $\sum_{i=1}^{|\mathcal{D}_L|} x_{ij} \cdot \Pr(D_{L,i}) = \Pr(D_{H,j})\ \forall j$ such that $1 \leq j \leq |\mathcal{D}_M|$.
Finally, we need to show that if $x_{ij} = y_{ij} \cdot \frac{\Pr(D_{H,j})}{\Pr(D_{L,i})}$, then $0 \leq x_{ij} \leq 1$. Note that $y_{ij} = 0$ when $D_{M,j} \cap U_L \nsubseteq D_{L,i}$, so $x_{ij} = 0$ when $D_{H,j} \cap U_L \nsubseteq D_{L,i}$. When $D_{H,j} \cap U_L \subseteq D_{L,i}$, $0 < y_{ij} \leq 1$. Also note that when $D_{H,j} \cap U_L \subseteq D_{L,i}$, $D_{H,j}$ and $D_{L,i}$ have some k number of words in common and the remaining $d-k$ words in $D_{H,j}$ are only "high" words. Therefore, the remaining $d-k$ words in $D_{H,j}$ are all lower frequency words than the remaining $d-k$ words in $D_{L,i}$. By repeatedly applying Lemma 17, we get that $\Pr(D_{H,j}) < \Pr(D_{L,i})$, and therefore, $0 \leq x_{ij} \leq 1$. □

Lemma 15. For every $D_{1,L}$, where $D_{1,L}$ is a dictionary sampled with respect to the L effort level, and for every g that satisfies the property that $D_{1,L}$ is mapped to a dictionary in $D_{1,H} \in \mathcal{D}_H$ such that $D_{1,H} \cap U_L \subseteq D_{1,L}$ and when both players play increasing frequency in the second stage and for all utility functions that satisfy rare-words preferences and $\alpha_k \geq \frac{\Pr(w_{n-k} \in D_H)}{\Pr(w_{n-k+1} \in D_H)}$ for all k, we have that:

$$\sum_{D_{2,H}} \Pr(D_{2,H}) \cdot u(s_1^{\uparrow}(g(D_{1,L})), s_2^{\uparrow}(D_{2,H})) \geq \sum_{D_{2,H}} \Pr(D_{2,H}) \cdot u(s_1^{\uparrow}(D_{1,L}), s_2^{\uparrow}(D_{2,H}))\ \forall D_{1,L} \tag{A.2}$$

In addition, the inequality is strict when $g(D_{1,L}) \neq D_{1,L}$.

Proof. Lemma 14 tells us that each D_L is mapped to a D_H that dominates it. Therefore, it suffices to show:

$$\sum_{D_{2,H}} \Pr(D_{2,H}) u(s_1^{\uparrow}(D'), s_2^{\uparrow}(D_{2,H})) \geq \sum_{D_{2,H}} \Pr(D_{2,H}) u(s_1^{\uparrow}(D), s_2^{\uparrow}(D_{2,H}))$$

for $D' = \{w'_d, w'_{d-1}, ..., w'_1\}$ and $D = \{w_d, w_{d-1}, ..., w_1\}$ (sorted in order of increasing frequency), where $w'_j = w_j$ for all $j \neq i$ and $f(w'_i) < f(w_i)$.
We can write $\sum_{D_{2,H}} \Pr(D_{2,H}) u(s_1^{\uparrow}(D'), s_2^{\uparrow}(D_{2,H}))$ as $\sum_{j=d}^{1} t'_j \cdot v(w_j)$, where $t'_d = \Pr(w_d \in D_{2,H})$, $t'_{d-1} = \Pr(w_d \notin D_{2,H} \cap w_{d-1} \in D_{2,H})$, ..., $t'_1 = \Pr(w_d \notin D_{2,H} \cap ... \cap w_2 \notin D_{2,H} \cap w_1 \in D_{2,H})$

Likewise $\sum_{D_{2,H}} \Pr(D_{2,H}) u(s_1^{\uparrow}(D'), s_2^{\uparrow}(D_{2,H})) = \sum_{j=d}^{1} t_j \cdot v(w_j)$ for analogous definition of the t_j's. Since D and D' differ in only the i^{th} coordinate, $t'_j = t_j$ for all $d \geq j \geq i+1$. Therefore:

$$\sum_{D_{2,H}} \Pr(D_{2,H}) u(s_1^{\uparrow}(D'), s_2^{\uparrow}(D_{2,H})) - \sum_{D_{2,H}} \Pr(D_{2,H}) u(s_1^{\uparrow}(D), s_2^{\uparrow}(D_{2,H}))$$
$$= \sum_{j=i}^{1} t'_j \cdot v(w_j) - \sum_{j=i}^{1} t_j \cdot v(w_j)$$

$t'_j > t_j$ for all $i-1 \geq j \geq 1$, so $t'_j \cdot v(w_j) > t_j \cdot v(w_j)$ for all $i-1 \geq j \geq 1$ and therefore, $\sum_{j=i-1}^{1} t'_j \cdot v(w_j) - \sum_{j=i-1}^{1} t_j \cdot v(w_j) > 0$.
Finally it suffices to show that $t'_i \cdot v(w'_i) \geq t_i \cdot v(w_i)$. We know from the statement of the theorem that $\frac{v(w'_i)}{v(w_i)} \geq \frac{\Pr(w_i \in D_H)}{\Pr(w'_i \in D_H)} = \frac{t_i}{t'_i}$, which gives us the desired result. □

Lemma 16. Given that players are playing words in order of increasing frequency,

$$\sum_{D_{1,H}} \Pr(D_{1,H}|H) \sum_{D_{2,H}} \Pr(D_{2,H}|H) \cdot u(s_1^{\uparrow}(D_{1,H}), s_2^{\uparrow}(D_{2,H})) >$$
$$\sum_{D_{1,L}} \Pr(D_{1,L}|L) \sum_{D_{2,H}} \Pr(D_{2,H}|H) \cdot u(s_1^{\uparrow}(D_{1,H}), s_2^{\uparrow}(D_{2,H})) \tag{A.3}$$

for all u that satisfy rare-words preferences and $\alpha_k \geq \frac{\Pr(w_{n-k} \in D_H)}{\Pr(w_{n-k+1} \in D_H)}$ for all k.

Proof. From Lemma 13, we know that:

$$\sum_{D_{1,H}} \Pr(D_{1,H}|H) \sum_{D_2} \Pr(D_{2,H}|H) \cdot u(s_1^{\uparrow}(D_{1,H}), s_2^{\uparrow}(D_{2,H})) =$$
$$\sum_{D_{1,L}} [\sum_{D_{1,H}} \Pr(D_{1,H}|H) \Pr(g(D_{1,L}) = D_{1,H})] \sum_{D_2} \Pr(D_{2,H}|H) u(s_1^{\uparrow}(D_{1,H}), s_2^{\uparrow}(D_{2,H}))$$

This is equivalent to writing:

$$\sum_{D_{1,H}} \Pr(D_{1,H}|H) \sum_{D_{2,H}} \Pr(D_{2,H}|H) \cdot u(s_1^{\uparrow}(D_{1,H}), s_2^{\uparrow}(D_{2,H}))$$
$$= \sum_{D_{1,H}} [\sum_{D_{1,L}} \Pr(D_{1,L}|L) \Pr(g(D_{1,L}) = D_{1,H}) \sum_{D_{2,H}} \Pr(D_{2,H}|H) u(s_1^{\uparrow}(g(D_{1,L})), s_2^{\uparrow}(D_{2,H}))]$$
$$> \sum_{D_{1,H}} [\sum_{D_{1,L}} \Pr(D_{1,L}|L) \Pr(g(D_{1,L}) = D_{1,H}) \sum_{D_{2,H}} \Pr(D_{2,H}|H) u(s_1^{\uparrow}(D_{1,L}), s_2^{\uparrow}(D_{2,H}))]$$
$$= \sum_{D_{1,L}} \Pr(D_{1,L}|H) \cdot \sum_{D_{2,H}} \Pr(D_{2,H}|H) \cdot u(s_1^{\uparrow}(D_{1,L}), s_2^{\uparrow}(D_{2,H}))$$

where the inequality follows from Lemma 15. □

Proposition 8. There exists a distribution over U for which $((H, s_1), (H, s_2))$ cannot be a Bayesian-Nash equilibrium of the ESP game for any pair of consistent second-stage strategies s_1, s_2 and for any utility function satisfying match-early preferences.

Proof. Similar to the other proofs of this nature, we map each high dictionary to a set of low dictionaries, using the randomized mapping from Section 4, that do at least as well as the high dictionary in expectation across the distribution of opponent dictionaries, with at least one high dictionary mapped to a low dictionary with non-zero probability that does strictly better than it. For the purpose of this proof we use the following distribution over U, each of the low words is sampled with probability $\frac{1-\epsilon}{|U_L|}$ and each of the high words is sampled with probability $\frac{\epsilon}{|U_H|-|U_L|}$. Note that under this distribution, the probability of sampling a dictionary consisting of "low effort words" only, is great than $d! \cdot (\frac{1-\epsilon}{|U_L|})^d$. The probability of sampling a dictionary that doesn't consist of entirely low effort words is $1-(1-\epsilon)^d$. We note that as $\epsilon \to 0$, $d! \cdot (\frac{1-\epsilon}{|U_L|})^d \to d! \cdot (\frac{1}{|U_L|})^d > 0$ and $1-(1-\epsilon)^d \to 0$. Thus there exists a value of ϵ such that $d! \cdot (\frac{1-\epsilon}{|U_L|})^d > 1-(1-\epsilon)^d$. We note that since the distribution described is uniform over the low effort words, if player 1 chooses L effort, his second-stage strategy is given by s_2, applied to the set of L words (from Lemma 4).

It suffices to show that: $\sum_{D_2} \Pr(D_2) \cdot I(l_{\leq k}(s_2(h(D_{1,H}))) \cap l_{\leq k}(s_2(D_2))) \geq \sum_{D_2} \Pr(D_2) \cdot I(l_{\leq k}(s_2(D_{1,H})) \cap l_{\leq k}(s_2(D_2))$ for all k and all $D_{1,H}$ with the inequality strict for some $D_{1,H}$ and some value of k. We know from the mapping that each dictionary $D_{1,H}$ is mapped to all dictionaries $D_{1,L}$ that satisfy the property $D_{1,H} \cap U_L \subseteq D_{1,L}$. Consider the prefix $l_{\leq k}(s_2(D_{1,H}))$ and the prefix $l_{\leq k}(s_2(D_{1,L}))$ where $D_{1,H}$ is mapped to $D_{1,L}$ under h. If $l_{\leq k}(s_2(D_{1,H})) \neq l_{\leq k}(s_2(h(D_{1,H})))$, $l_{\leq k}(s_2(D_{1,H}))$ contains a set of "high effort" words and/or it contains a set of "low effort words". If it contains some "low effort words" that are not in $l_{\leq k}(s_2(D_{1,L}))$, these words are in $D_{1,L}$ but are lower priority words than all of the words in $l_{\leq k}(s_2(D_{1,L}))$. It suffices to show $\sum_{D_2} \Pr(D_2) \cdot I(l_{\leq k}(s_2(D)) \cap l_{\leq k}(s_2(D_2)) \geq \sum_{D_2} \Pr(D_2) \cdot I(l_{\leq k}(s_2(D')) \cap l_{\leq k}(s_2(D_2))$ (eq. *) for a pair of dictionaries D and D' where $l_{\leq k}(s_2(D')) = l_{\leq k}(s_2(D)) - \{w_i\} + \{w_j\}$ and either $w_i, w_j \in U_L$ with $w_i \succ w_j$ under s_2 or $w_i \in U_L$ and $w_j \in U_H$. First we handle the case where $w_i, w_j \in U_L$ with $w_i \succ w_j$ under s_2. Now we handle the case where $w_i \in U_L$ and $w_j \in U_H$. Since $w_i \succ w_j$ under s_2 and $f(w_i) = f(w_j)$, $Pr(w_i \in l_{\leq k}(s_2(D_2)) > Pr(w_j \in l_{\leq k}(s_2(D_2))$, by Lemma 4. Therefore, we know that (eq. *) is satisfied for this case and the inequality is strict. Since w_i is in the top k words of D, there exists at least one dictionary D_2 with w_i in the top k words. This dictionary occurs with probability greater than any high effort word occurs in a dictionary D_2. Therefore, we know that (eq. *) is satisfied for this case and the inequality is strict. Since we know that there exists at least one value of $D_{1,H}$ such that $h(D_{1,H}) \neq D_{1,H}$, there exists a value of k and $D_{1,H}$ such that the inequality is strict. Therefore, playing (L, s_2) is a strict ordinal best response to (H, s_2). □

Lemma 18. $\frac{1-(1-\frac{1}{n^s Z})^d}{1-(1-\frac{1}{m^s Z})^d} \leq \frac{m^s}{n^s}$ *for all* $Z > 0$ *and all integers* m, n, d *such that* $m > n \geq 1$ *and* $d \geq 1$ *and all* $s \geq 1$.

Proof. It suffices to show that the above statement holds for $d = 1$ and that $\frac{1-(1-\frac{1}{n^s Z})^d}{1-(1-\frac{1}{m^s Z})^d} < \frac{1-(1-\frac{1}{n^s Z})}{1-(1-\frac{1}{m^s Z})}$. Observe that $\frac{1-(1-\frac{1}{n^s Z})}{1-(1-\frac{1}{m^s Z})} = \frac{\frac{1}{n^s Z}}{\frac{1}{m^s Z}} = \frac{m^s}{n^s}$. Hence the statement is true for $d = 1$.

Now consider $\frac{1-(1-\frac{1}{n^s Z})^d}{1-(1-\frac{1}{m^s Z})^d}$

This expression is the same as $\frac{(1-(1-\frac{1}{n^s Z}))\cdot(1+(1-\frac{1}{n^s Z})+\ldots+(1-\frac{1}{n^s Z})^{d-1})}{(1-(1-\frac{1}{m^s Z}))\cdot(1+(1-\frac{1}{m^s Z})+\ldots+(1-\frac{1}{m^s Z})^{d-1})}$

Since $(1-\frac{1}{n^s Z}) < (1-\frac{1}{m^s Z})$, we have that $\frac{1+(1-\frac{1}{n^s Z})+\ldots+(1-\frac{1}{n^s Z})^{d-1}}{1+(1-\frac{1}{m^s Z})+\ldots+(1-\frac{1}{m^s Z})^{d-1}} < 1$

$\frac{1-(1-\frac{1}{n^s Z})^d}{1-(1-\frac{1}{m^s Z})^d} < \frac{1-(1-\frac{1}{n^s Z})}{1-(1-\frac{1}{m^s Z})}$. □

Lemma 19. *For every pair of D_H and D_L such that $D_H \cap U_L \subseteq D_L$, when player 2 is playing L effort and both players play increasing frequency in the second stage, we have that:*

$$\sum_{D_2} \Pr(D_2) \cdot I(g_w(s_1^{\uparrow}(D_L), s_2^{\uparrow}(D_2)) \in \{w_n, ..., w_{n-k+1}\}) \geq \sum_{D_2} \Pr(D_2) \cdot I(g_w(s_1^{\uparrow}(D_H), s_2^{\uparrow}(D_2)) \in \{w_n, ..., w_{n-k+1}\}) \ \forall k \tag{A.4}$$

When $D_H \cap D_L = \emptyset$, this inequality is strict for all k.

Proof. Since player 2 is playing L effort, $\sum_{D_2} \Pr(D_2) \cdot I(g_w(s_1^{\uparrow}(D_L), s_2^{\uparrow}(D_2)) \in \{w_n, ..., w_{n-k+1}\}) = \sum_{D_2} \Pr(D_2) \cdot I(g_w(s_1^{\uparrow}(D_H), s_2^{\uparrow}(D_2)) \in \{w_n, ..., w_{n-k+1}\}) = 0$ for all $1 \leq k \leq |U_H| - |U_L|$.
Now consider $|U_H| - |U_L| \leq k \leq |U_H|$.
Let $A = D_L \cap \{w_{|U_L|}, w_{|U_L|-1}, ..., w_{|U_L|-k+1}\}$ and $B = D_H \cap \{w_{|U_L|}, w_{|U_L|-1}, ..., w_{|U_L|-k+1}\}$. Therefore, we can write:
$\sum_{D_2} \Pr(D_2) \cdot I(g_w(s_1^{\uparrow}(D_L), s_2^{\uparrow}(D_2)) \in \{w_n, ..., w_{n-k+1}\}) = \sum_{D_2} \Pr(D_2) \cdot I(A \cap s_2^{\uparrow}(D_2))$ and
$\sum_{D_2} \Pr(D_2) \cdot I(g_w(s_1^{\uparrow}(D_H), s_2^{\uparrow}(D_2)) \in \{w_n, ..., w_{n-k+1}\}) = \sum_{D_2} \Pr(D_2) \cdot I(B \cap s_2^{\uparrow}(D_2))$.
Since $D_H \cap U_L \subseteq D_L$, we get $B \subseteq A$. Therefore, if $I(B \cap s_2^{\uparrow}(D_2)) = 1$, then $I(A \cap s_2^{\downarrow}(D_2)) = 1$, for all $D_2 \in \mathcal{D}_L$. Thus, $\sum_{D_2} \Pr(D_2) \cdot I(A \cap s_2^{\uparrow}(D_2)) \geq \sum_{D_2} \Pr(D_2) \cdot I(B \cap s_2^{\uparrow}(D_2))$. If $D_H \cap D_L = \emptyset$, then there exists $D_2 \in \mathcal{D}_L$ such that $A \cap D_2 \neq \emptyset$ and $B \cap D_2 = \emptyset$. Thus in this case, we get that: $\sum_{D_2} \Pr(D_2) \cdot I(A \cap s_2^{\uparrow}(D_2)) > \sum_{D_2} \Pr(D_2) \cdot I(B \cap s_2^{\uparrow}(D_2))$, which gives us the desired result. □