Article

Entropy and the Predictability of Online Life

Roberta Sinatra [1] **and Michael Szell** [2,*]

[1] Center for Complex Networks Research and Physics Department, Northeastern University, 110 Forsyth Street, Boston, MA 02115, USA; E-Mail: r.sinatra@neu.edu

[2] Senseable City Laboratory, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

* Author to whom correspondence should be addressed; E-Mail: mszell@mit.edu; Tel.: +1-617-324-4474; Fax: +1-617-258-8081.

Abstract: Using mobile phone records and information theory measures, our daily lives have been recently shown to follow strict statistical regularities, and our movement patterns are, to a large extent, predictable. Here, we apply entropy and predictability measures to two datasets of the behavioral actions and the mobility of a large number of players in the virtual universe of a massive multiplayer online game. We find that movements in virtual human lives follow the same high levels of predictability as offline mobility, where future movements can, to some extent, be predicted well if the temporal correlations of visited places are accounted for. Time series of behavioral actions show similar high levels of predictability, even when temporal correlations are neglected. Entropy conditional on specific behavioral actions reveals that in terms of predictability, negative behavior has a wider variety than positive actions. The actions that contain the information to best predict an individual's subsequent action are negative, such as attacks or enemy markings, while the positive actions of friendship marking, trade and communication contain the least amount of predictive information. These observations show that predicting behavioral actions requires less information than predicting the mobility patterns of humans for which the additional knowledge of past visited locations is crucial and that the type and sign of a social relation has an essential impact on the ability to determine future behavior.

1. Introduction

Capturing the regularities of our daily lives and the occasional deviations from the steady diurnal patterns has traditionally eluded an all-encompassing approach, due to tremendous efforts in monitoring detailed human activities over long times and the bias in behavior caused by obtrusive methods of observation [1]. However, the recent ability to address questions in social science by using huge datasets that have emerged over the past decades as a result of digitalization has opened previously unimaginable ways of conducting research in the field [2].

On the one hand, these new datasets give a highly detailed protocol of our ordinary lives, for example, in the form of mobile phone data, which enables a deeper understanding of the regularities in our mobility patterns [3,4], and how the regularities in human behavior are reflected in the geographic regions that emerge from our interactions [5]. On the other hand, from a previous point of view, "extraordinary" new forms of human behavior can now be observed online, where the full set of all actions performed in the system is typically available for study, spanning an even deeper level of detail. Online social networking services, such as Twitter or Facebook, or discussion forums allow new insights into the rhythms of social actions and interactions, as expressed online [6–9], and how these interactions relate to the underlying offline events [10]. An even richer insight can be gained into human-led lives that unfold *entirely* in artificial online environments, such as in persistent, massive multiplayer online games, where human-controlled characters spend their whole virtual lives within an online world interacting with other characters [11]. The playing of online games is one of the most wide-spread forms of collective human behavior in the world; the "massive multiplayer" aspect allows one to not only study single individuals, but also collective behavioral phenomena that typically emerge in complex social systems [12]. Here, data can be available of all actions, decisions and interactions between many thousands of individuals over long time spans [13], allowing understanding of the structure and evolution of socio-economic networks [14–16], mobility [17] or the emergence of good conduct [18] and elite structures [19] in large social systems.

Going hand in hand with the new availability of large-scale, longitudinal behavioral datasets of various kinds, well-known methods from the mathematical and physical sciences, especially statistical physics and information theory [20–22], have been extended and/or re-applied successfully in this context. In particular, principal component analysis and the concept of "eigenbehavior" has been used to quantify behavioral regularities and to predict future activities in the daily lives of a group of 100 subjects [23]. Similarly, information theory measures provide an adequate quantification between uniform distributions (maximal entropy) and maximally uneven distributions of states (minimal entropy), which, in the case of human behavior, can inform us about the extent of uniformity and, thus, predictability in our activity patterns. The concept of entropy has been applied specifically to assess the predictability of mobility patterns [24,25], of economic behavior [26], the order of human-built structures, such as urban street networks [27], or the complexity of online chatting behavior [9,28,29]. Further, a theoretical framework for non-extensive entropies has been recently developed that might be well applicable to complex systems [30,31].

2. Behavior and Mobility Data of Human Players in the Online World, Pardus

Here, our goal is to apply classical entropy measures to study the patterns of various kinds of behavior in a single, closed socio-economic system, as generated by thousands of users in the online game "Pardus" [13], to provide an insight into the regularity of life in online worlds and, eventually, to draw possible conclusions on how humans lead their offline lives.

2.1. The Online World, Pardus

The online world Pardus, www.pardus.at, is a browser-based, massive multiplayer online game open to the public for over nine years. Over 400,000 users have registered to play so far. The game features three independent, persistent game universes, which had a defined starting time, but no scheduled end. There are no predefined goals in the game: many aspects of social life within Pardus are self-organized, for example, the emergence of social groups (alliances) and the politics between them. Players are engaged in a multitude of social activities, *i.e.*, chatting, cultivating friendships, building up alliances, but also negative interactions, such as destructive attacks, and economic activities, such as producing commodities in factories and selling them to other players. We focus on the "Artemis" game universe, in which we recorded player actions over the first 1,238 consecutive days of the universe's existence. Communication between any two players can take place directly, by using a one-to-one, e-mail-like private messaging system. We focus on one-to-one interactions between players only and discard indirect interactions, such as, e.g., participation in chats or forums. There are global interactions, *i.e.*, interactions that can be performed independently of the spatial position of players in the game universe, which are communication, setting and removing friendship or enemy links, or placing a bounty on another player. The actions of trade or attack, however, need players to meet in space. All data used in this study are fully anonymized.

2.2. Mobility Time Series

For studying regularities in mobility patterns, we use the same dataset of Pardus player movements that has been used in [17]. A universe in Pardus can be represented as a network with 400 nodes, called sectors, and 1,160 links. Each sector is like a city, where players can have social relations or entertain economic activities. Typically, sectors adjacent on the universe map, as well as a few far-apart sectors, are interconnected by links that allow players to move from sector to sector. At any point in time, each sector is usually attended by a large number of players. The universe network has a large diameter of 27, which means that, on average, players have to move through a non-negligible number of sectors to traverse the universe. Due to a limited pool of actions that players can spend on movement, traveling large distances can take a player several days. Using this dataset, we previously studied the statistical movement patterns of players and found that locations are visited in a specific order, leading to strong long-term memory effects [17]. In detail, we extract player mobility data from day 200 to day 1,200 of the universe's existence. We discard the first 200 days, because social networks between players of Pardus have shown aging effects in the beginning of the universe [15]. To make sure we only consider active players, we select all who exist in the game between the days 200 and 1,200, yielding 1,458

players active over a time-period of 1,000 days. The sector IDs of these players, *i.e.*, their positions on the universe network's nodes, are logged every day at 05:35 GMT.

2.3. Behavioral Action Time Series

For studying regularities in behavioral time series, we use the same dataset of Pardus player actions that has been used in [18]. Players can express their sympathy (distrust) toward other players by establishing so-called friendship (enmity) links. These links are only seen by the player marking another as a friend (enemy) and the respective recipient of that link. For more details on the game, see [15]. We consider eight different actions every player can execute at any time. These are communication (C), trade (T), setting a friendship link (F), removing an enemy link (forgiving) (X), attack (A), placing a bounty on another player (punishment) (B), removing a friendship link (D) and setting an enemy link (E). While C, T, F and X can be associated with positive actions, A, B, D and E are hostile or negative actions. We classify communication as positive, because only a negligible part of communication takes place between enemies [15]. Following a previous formalism [15], we say that positive actions have a positive sign, and negative actions have a negative sign. The alphabet, $\mathcal{X}$, of all possible dyadic actions happening in each player's life therefore spans 16 letters: eight possible performed actions (four negative, four positive) and eight possible received actions (four negative, four positive). We denote received actions with the suffix, r, e.g., A_r for a received attack. Due to the heterogeneous activity patterns of players, we operate in action-time rather than in actual time; for example, indices of t and $t-1$ denote that two actions were subsequent, regardless of whether the actual time difference was seconds or weeks [18]. From all sequences of all 34,055 Artemis players who performed or received an action at least once within 1,238 days, we removed players with a life history of less than 1,000 actions, leading to the set of the most active 1,758 players that are considered throughout this work.

3. Entropy and Predictability Measures

To study the regularity and predictability of behavior from the discrete time series, we use three entropy measures. Following [24], we call the binary logarithm of the number of distinct states, N_i, of a player, i, the *random entropy*:

$$S_i^{\mathrm{rand}} = \log_2 N_i \tag{1}$$

In the case of mobility, "states" refer to the 400 possible sectors in the universe visitable at a given point in time by a player. The maximal possible random entropy is $S^{\mathrm{rand}} = \log_2 400 \approx 8.6$, reached when all sectors are visited at least once. In the case of behavioral actions, a state can be one of the 16 possible action or received action types; here, the maximum possible random entropy is $S^{\mathrm{rand}} = \log_2 16 = 4$.

The Shannon entropy, S_i^{unc}, of a player, i, is defined as:

$$S_i^{\mathrm{unc}} = -\sum_{x \in \mathcal{X}_i} p_i(x) \log_2 p_i(x) \tag{2}$$

where $p_i(x)$ is the measured probability over the respective time span that player i has occupied a state, x, and $\mathcal{X}_i$ is the ensemble of the N_i distinct states. In this context, we call the Shannon entropy the *temporal-uncorrelated entropy*, because it captures the entropy when the temporal order of states is

ignored [24]. The random and temporal-uncorrelated entropies are equal, $S_i^{\mathrm{rand}} = S_i^{\mathrm{unc}}$, if all of the N_i distinct states, x, were occupied with uniform probability $p_i(x) = 1/N_i$ by the player, i. For mobility, the occupation of a single sector over the whole time span of 1,000 days would result in the smallest possible random and temporal-uncorrelated entropy of $S^{\mathrm{rand}} = S^{\mathrm{unc}} = 0$.

Finally, we make use of the conditional entropy S_i^{cond} of a player, i, capturing the entropy conditional on temporal short-term correlations over one previous state in the time series,

$$S_i^{\mathrm{cond}} = - \sum_{x_t \in \mathcal{X}_i} \sum_{x_{t-1} \in \mathcal{X}_i} p_i(x_{t-1}, x_t) \log_2 p_i(x_t | x_{t-1}) \tag{3}$$

with $p_i(x_{t-1}, x_t)$ being the probability of occurrence of the pair of subsequent states, x_{t-1} and x_t, $p_i(x_t | x_{t-1}) = p_i(x_{t-1}, x_t)/p(x_{t-1})$, the probability of the state, x_t, at time t given a preceding state, x_{t-1}. The conditional and temporal-uncorrelated entropies are equal, $S_i^{\mathrm{cond}} = S_i^{\mathrm{unc}}$, if there are no temporal correlations.

It is easy to show that we have $S^{\mathrm{cond}} \le S^{\mathrm{unc}} \le S^{\mathrm{rand}}$ for each user [32]. The differences in these two inequalities quantify the effects of short-term temporal correlations and the uniformity of the occupation distribution, respectively. To assess the predictability of specific states or of classes of states, we also define the conditional entropy for the set of states, $\mathcal{Z}$,

$$S_i^{\mathrm{cond}}(\mathcal{Z}) = - \sum_{x_t \in \mathcal{X}_i} \sum_{x_{t-1} \in \mathcal{X}_i \cap \mathcal{Z}} p_i(x_{t-1}, x_t) \log_2 p_i(x_t | x_{t-1}) \tag{4}$$

which is the conditional entropy given that the previous state belonged to $\mathcal{Z}$, where $\mathcal{Z}$ can be fixed as any subset of all the possible states, $\mathcal{X}$. Notice that $S_i^{\mathrm{cond}} \equiv S_i^{\mathrm{cond}}(\mathcal{Z}) + S_i^{\mathrm{cond}}(\mathcal{X}_i \setminus \mathcal{Z})$.

Complementary to entropy measures of information content or *unpredictability* are measures of *predictability* that denote in a percent value how likely an appropriate predictive algorithm could foresee an individual's future behavior [24]. The predictability, $\Pi_i^{\bullet}$, of an individual i is bounded above by:

$$S_i^{\bullet} = H(\Pi_i^{\bullet}) + (1 - \Pi_i^{\bullet}) \log_2(N_i - 1) \tag{5}$$

with the binary entropy function:

$$H(\Pi_i^{\bullet}) = \Pi_i^{\bullet} \log_2(\Pi_i^{\bullet}) - (1 - \Pi_i^{\bullet}) \log_2(1 - \Pi_i^{\bullet}) \tag{6}$$

where $\bullet$ is a placeholder for any of the types, rand, unc or cond. Unlike the measure of entropy, which is well established, the application of this predictability measure to practical problems is relatively recent. It is based on the idea that predictability is related to the error probability in guessing the outcome of a discrete random variable [33]. The upper bound given in Equation (5) comes from Fano's inequality [32,34]. For a detailed discussion on this bound and on possible lower bounds, see [24,33].

For being able to study in more detail the effects of memory in the system [35,36], we generalize the conditional entropy:

$$S_i^{\mathrm{cond},k} = - \sum_{x_t \in \mathcal{X}_i} \cdots \sum_{x_{t-k} \in \mathcal{X}_i} p_i(x_{t-k}, \ldots, x_t) \log_2 p_i(x_t | x_{t-k}, \ldots, x_{t-1}) \tag{7}$$

where k is an integer denoting the memory window. Note that $S_i^{\mathrm{cond},1} \equiv S_i^{\mathrm{cond}}$ and that we can identify $S_i^{\mathrm{cond},0}$ with S_i^{unc}. It follows from Fano's inequality [34] that $S_i^{\mathrm{cond},1} \ge S_i^{\mathrm{cond},2} \ge S_i^{\mathrm{cond},3} \ge \cdots$.

The differences between subsequent values in this chain inform us about the gain of predictability when we increase the memory window one by one. If such a difference starts becoming negligible from a particular level, k to $k+1$, it means that the system does not exhibit relevant memory effects beyond a window of k steps. If this level is at $k = 0$, the events are uncorrelated; if at $k = 1$, the system is Markovian, otherwise, it is non-Markovian.

4. Results and Discussion

4.1. Predictability in Mobility

We applied all entropy and predictability measures to the mobility time series, Figure 1a,b, respectively. Results show almost identical predictability behavior for humans in our online world as for the mobility of humans in geographic space [22,24]. The distributions for S^{unc} and S^{rand} are both qualitatively and quantitatively matching, showing that also online, movements of human avatars have the same highly predictable patterns when temporal correlations are accounted for, but are mostly unpredictable when the order of visitations is ignored. In particular, also here, S^{rand} peaks around six, indicating that an individual who chooses her next location randomly could be found, on average, in any of $2^{S^{\text{rand}}} \approx 64$ locations, which is a substantial part of the 400 possible sectors. The contrasting peak of S^{cond} below two shows that the actual uncertainty of a typical player's location is not 64, but rather, less than $2^2 = 4$ sectors. The conditional entropy, S^{cond}, is not directly comparable to the actual entropy, S, in [24], but shows the same tendency in that temporal correlations are substantial, even if just having a memory of one. However, for re-creating the statistical features of mobility thoroughly, longer memory is needed [17]. The peak of Π^{cond} around 0.9 means that only in around 10% of cases does a player choose her location in a manner that appears to be random, but in 90% of the cases, we can hope to predict her whereabouts with an appropriate predictive algorithm. This high predictability stands in contrast to the moderately predictive case given by Π^{unc} peaking around 0.5 and the highly unpredictive case of Π^{rand} peaking narrowly and close to zero.

Figure 1. The distribution of (**a**) entropy and (**b**) the predictability measures of the mobility of the Pardus players. Both are almost identical to the mobility of humans in geographic space [24]: Each considered entropy measure improves predictability substantially, from considering the uniformity of occupation to additionally short-term temporal correlations.
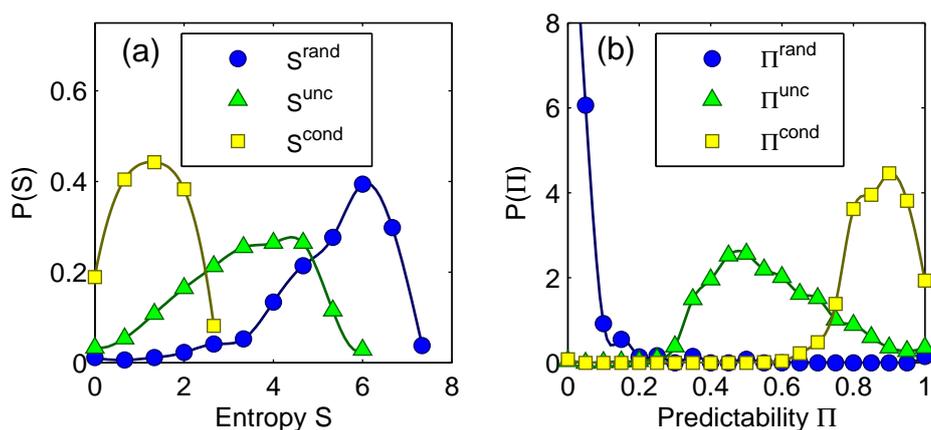
4.2. Predictability in Behavioral Actions

A similar picture to mobility arises for behavioral actions. Figure 2a,b, respectively, report the entropy and predictability distributions of all 16 types of actions and received actions. Here, S^{rand} is peaked at four, showing that most players are making full use of their behavioral possibilities of $16 = 2^4$ action and received action types in the course of their online lives. However, the sharp drop to the distribution of S^{unc}, which peaks around two, shows that, in practice, most of these actions and received actions are focused on around $2^2 = 4$ action or received action types only. The even narrower curve of S^{cond}, which peaks around 1.5, with a corresponding peak of Π^{cond} at 0.8, demonstrates that the conditional information allows us to predict 80% of actions. This is only slightly more than the 73% prediction rate peak from Π^{unc}; however, Π^{cond} is distributed more widely. In conclusion, the predictability gained from considering the uniformity of occupation is much larger than the predictability gained from also considering Markovian temporal correlations, as opposed to the case of mobility where temporal correlations add substantial predictive value.

Figure 2. Distribution of (**a**) entropy and (**b**) predictability measures of the behavioral actions of the Pardus players. As in the case of mobility, behavioral actions are highly regular and predictable. However, the predictability gained from considering the uniformity of occupation is much larger than the predictability gained from also considering temporal correlations.
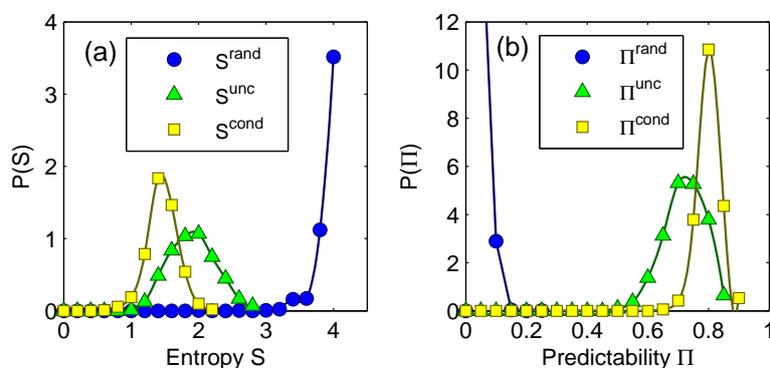
One previous key observation on Pardus players is the fundamental structural and dynamic difference between positive and negative action types and their interaction networks [13–15,18]. To see if this difference is also apparent in the extent of predictability, we plotted the distribution of the conditional entropy of the players given that the previous action or received action was positive/negative (Figure 3a), *i.e.*, the set $\mathcal{Z}$ in Equation (4) corresponds to $\mathcal{Z} = \{C, T, F, X, C_r, T_r, F_r, X_r\}$ or to $\mathcal{Z} = \{A, B, D, E, A_r, B_r, D_r, E_r\}$, respectively. We aim to understand whether the actions that follow positive actions are more predictable than those that follow negative actions. If the distributions were identical, the sign of an action would cause no difference in the predictability of the subsequent action. In fact, although both distributions peak around 0.55, showing that there is a moderate amount of predictive value gained from the information of an action's sign, the positive distribution is much more narrow than the negative one, implicating that there is a much wider range of negative behavior in terms of

predictability than positive behavior. This result suggests that "good" people are much alike, but "bad" persons behave badly in more various and, sometimes, more unpredictable ways.

Figure 3. The distribution of the conditional entropy measures of the behavioral actions of the Pardus players, given that the previous action belonged to a certain category. (**a**) Entropy given that the previous action or received action was positive/negative. The positive and negative distributions have their maxima both around 0.55, but the former is much more narrow than the latter one, showing that there is a much wider range of negative behavior in terms of predictability than positive behavior. (**b**) Entropy given that the previous action was performed/received. Both distributions peak very close to one, showing that the information of whether an action was performed or received does, in general, not have a high predictive value. The peak for the received actions is slightly closer to one than for the performed actions.
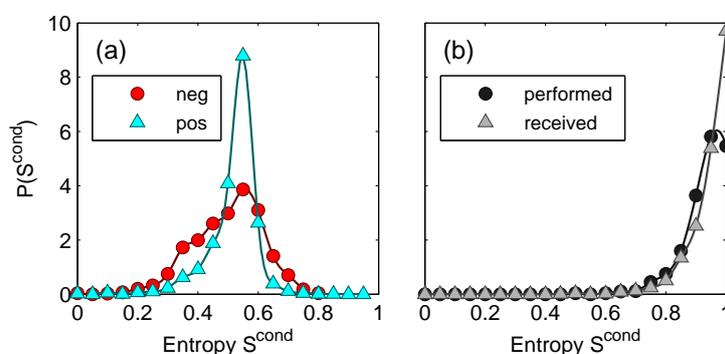

The conditional entropy for performed or received actions, *i.e.*, $\mathcal{Z} = \{C, T, F, X, A, B, D, E\}$ or $\mathcal{Z} = \{C_r, T_r, F_r, X_r, A_r, B_r, D_r, E_r\}$ in Equation (4), respectively, is peaked very narrowly and close to one for both cases and slightly more so for received actions; Figure 3b. This observation shows that the directionality of actions contains much less predictive information than the sign of an action.

We can further refine the conditional entropy measure by considering single actions as the condition, *i.e.*, where $\mathcal{Z}$ in Equation (4) is a singleton, to assess how much each action or received action type allows one to predict the subsequent action that is about to happen in a player's life. The conditional entropy of trade peaks around 1.3; the distribution of communication is more wide, peaking around six bits; Figure 4a. Distributions of received trades and communications are almost identical, only received communication is slightly more right-skewed than performed actions of communication. The reason why communication is associated with higher unpredictability might have to do with the game's action point system [15]: every action, except the action of communication, costs an amount of so-called action points for which every player has only a limited pool. Therefore, players are not limited in their communication behavior, but are so for trade, friendship markings, *etc*. The entropy distribution of friendship marking, F, Figure 4b, peaks around one bit and is, therefore, much less unpredictable. The entropy of enemy marking E peaks even closer to zero (Figure 4d); all of the actions related to enemy markings, E, E_r, X and X_r, show a bimodal distribution with an extra peak at zero, but this is clearly not the case for friendship markings F or F_r. This bimodality could hint towards two different kinds of effects that arise from enemy marking, where, for example, either the person who makes or removes the marking immediately predictably sends a message to the recipient in a fraction of cases,

or in the remaining fraction, this does not happen. Finally, the conditional entropy of received attacks, A_r, peaks around one, and performed attacks, A, are more wide peaking at a smaller value; Figure 4c. In all the distributions that deal with friendship or enemy markings, F, D, E and X, we observe a right-shift of peaks for received actions, meaning that a player's next action is more predictable given that a friend/enemy event happened to her, as opposed to when she performed such an action towards somebody else. For attacks, however, we see the opposite. It is unclear what causes this phenomenon or how relevant it is: we can only speculate that a received attack could have a possibly stronger emotional impact on a player and, therefore, a more adverse effect on the predictability on her next action, while this is *vice versa* for friendship/markings. Further, it is interesting to note that the removal of a friendship link has a similar pattern to the addition of an enmity link, suggesting that these two actions might be closely related, since they have a similar impact on future behavior. In general, however, the removal of a positive/negative tie cannot always be put on the same level as the addition of a negative/positive tie, as the reversed case of friendship addition and enemy removal shows.

Figure 4. The distribution of conditional entropy measures of the behavioral actions of the Pardus players, given that the previous action was of a certain type. (**a**) The distributions for performed and received communication events (C and C_r) and for performed and received trade events (T and T_r). Communication peaks around six bits, trade around 1.3 bits. Performed and received actions do not show substantial deviations here. (**b**) The distributions for performed and received friendship marking events (F and F_r) and for performed and received friendship removals (D and D_r). The curves peak around one or lower. (**c**) The distributions for performed and received attacks (A and A_r). The former curve peaks below one; the latter peaks around one and is narrower. (**d**) The distributions for performed and received enemy marking events (E and E_r) and for performed and received enemy removals (X and X_r). All the curves peak once around 0.6 and another time close to zero.
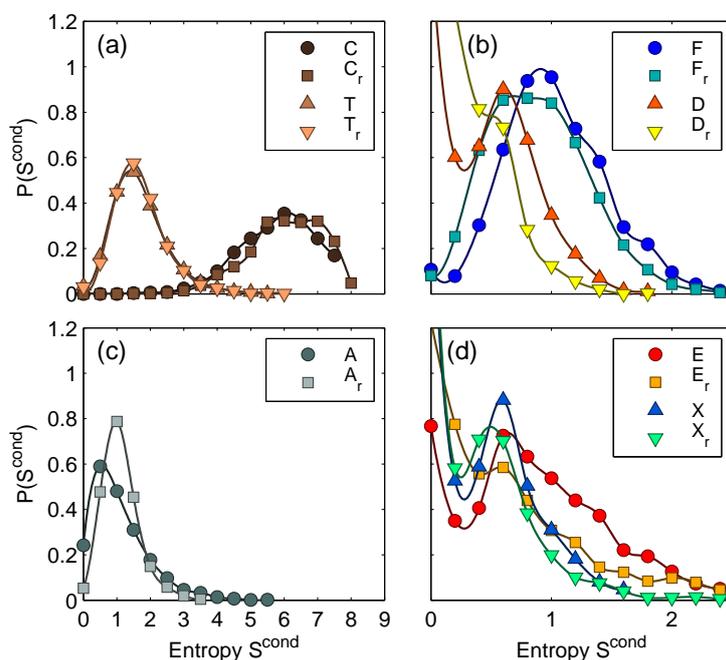
Finally, we are interested in assessing the memory dependence of the behavioral actions in the system [36], *i.e.*, the gain of predictability from conditional entropies with longer time windows, using the measures, $S^{\mathrm{cond},k}$, for increasing k. Unfortunately, in practice, these rely on the empirical probabilities, $p_i(x_{t-k}, \ldots, x_t)$, of all possible substrings $x_{t-k}, \ldots, x_t$ (see Equation (7)), which would lead to combinatorial explosion with our alphabet size of 16. For example, $k = 3$ would mean $16^3 = 4,096$ possible substrings of a length of three, many of which do not exist at all or are statistically not reliable to assess from a dataset of 1,758 players, each having performed up to a few thousand actions. Therefore, in the following, we used the simplified alphabet of a size of two of negative or positive actions, allowing feasible calculation of $S^{\mathrm{cond},k}$ up to $k = 5$. The distributions of these entropies are shown in Figure 5. The distributions converge quickly, showing only a small difference between $S^{\mathrm{cond},1}$ and $S^{\mathrm{cond},2}$ and almost no difference between higher order distributions. We quantify these differences via the Kullback–Leibler divergence between the distributions of the conditional entropy of subsequent memory levels, $S^{\mathrm{cond},k-1}$ and $S^{\mathrm{cond},k}$,

$$D(k) = D(S^{\mathrm{cond},k} || S^{\mathrm{cond},k-1}) = \sum_j S^{\mathrm{cond},k}(j) \log \frac{S^{\mathrm{cond},k}(j)}{S^{\mathrm{cond},k-1}(j)} \tag{8}$$

which provides the information gain for going from a memory of a length of $k-1$ to k [32,35]. A divergence of zero means that two distributions are identical. The first values from $D(2)$ to $D(5)$ read 0.0097, 0.0020, 0.0006 and 0.0005. For comparison, the Kullback–Leibler divergence between S^{unc} and S^{cond}, $D(1)$, yields the much higher value of 0.38, showing that the system is, to a large part, Markovian and that the predictability gained from higher-order correlations is negligible.

Figure 5. The convergence of the conditional entropy of the positive and negative behavioral actions of Pardus players with an increasing memory window. The difference between $S^{\mathrm{cond},1}$ and $S^{\mathrm{cond},2}$ is small, $D(2) = 0.0097$, showing that the system is almost Markovian. For higher memory windows, we have $D(3) = 0.0020$, $D(4) = 0.0006$ and $D(5) = 0.0005$, indicating almost identical distributions, which implies that there are practically no long-term correlations in the signs of behavioral actions.
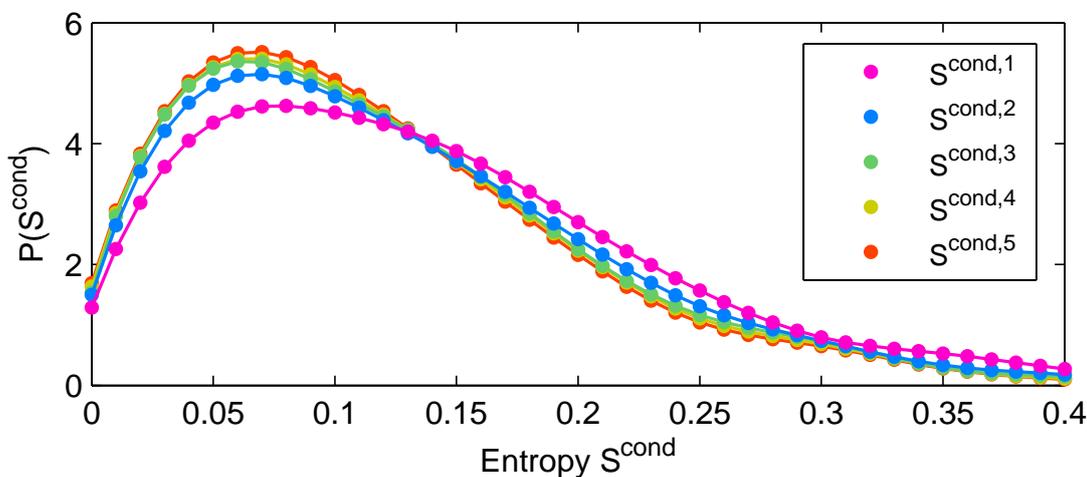
5. Conclusions

We applied three measures of entropy to two sets of time series of the behavioral actions and the movements of a large number of players in a virtual universe of a massive multiplayer online game. We found that movements in virtual human lives follow identical levels of predictability as offline mobility. This result reasserts previous observations on the similarities between the online and offline movements of humans [17] and is especially striking considering that in online worlds, individuals are not performing physical movements, but rather, navigate a virtual avatar.

Extending the approach to behavioral time series, also, here, we were able to provide evidence for high predictability. However, in this case, we found that due to weaker temporal correlations, there is hope to more easily predict behavioral actions than the temporally correlated mobility patterns of humans for which information about previously visited locations is required. Findings using entropy measures conditional on positive and negative actions suggest that "good" people are much alike, but "bad" persons behave badly in more various and, sometimes, more unpredictable ways. Actions containing the highest predictive information for an individual's next behavior are negative, such as attacks or enemy markings, while the positive actions of friendship marking, trade and communication contain the least amount of predictive information. However, we show that the system is, to a large part, Markovian and almost devoid of any higher order correlations when taking into account the sign of the action, showing that positive or negative behavior is not more predictable when a longer history of previous actions is accounted for.

The distributions of entropies and predictability found here is strikingly similar to distributions found for datasets of offline mobility [24], economic transactions [26], online conversations and online location check-ins [29], therefore suggesting a possible universality in the limitations of human behavior and its independence of the concrete medium or context. However, contrary to our result of little high-order correlations in behavior, a recent study has shown that the behavior of browsing web pages is, to a large extent, non-Markovian [36]. Non-extensive entropies have been recently developed that might be well applied for non-Markovian settings in complex social systems [30,31].

Our observations also provide additional evidence for the fundamental differences in positive and negative behavior that were previously found on dynamic [18] and structural [13–15] levels. Although previously large-scale evidence has confirmed in online human behavior a number of known or hypothesized behavioral phenomena of offline behavior, it is not immediately clear how asymmetries between positive and negative behavior in our, to some extent, artificial, online world can be translated to the offline world. Future research should aim to analyze positive and negative relationships and behaviors that happen in real-life societies and organizations [37], especially considering the multi-relational aspect of social organization [14,38]. Fine-grained datasets of socio-economic behavior, such as the one presented, offer the further possibility of going beyond observations and measurements, to study the mechanisms and origins of behavior in the view of collective phenomena [9].

6. Notes Added in Proof

During the redaction of this paper, we were made aware of a relevant study that applied the conditional entropy of signed messages to model growth of entropy in emotionally charged online dialogues [39].

Acknowledgments

Conflicts of Interest

Michael Szell is an associate of the company, Bayer & Szell OG, which is developing and maintaining the online game, Pardus, from which the data was collected.

References

1. Rosenthal, R. *Meta-Analytic Procedures for Social Research, Vol. 6.*; Sage: Newbury Park, CA, USA, 1991.
2. Lazer, D.; Pentland, A.; Adamic, L.; Aral, S.; Barabási, A.L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; *et al*. Computational social science. *Science* **2009**, *323*, 721–723.
3. González, M.; Hidalgo, C.; Barabási, A.L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779–782.
4. Schneider, C.M.; Belik, V.; Couronné, T.; Smoreda, Z.; González, M.C. Unravelling daily human mobility motifs. *J. R. Soc. Interface* **2013**, *10*, doi:10.1098/rsif.2013.0246.
5. Sobolevsky, S.; Szell, M.; Campari, R.; Couronné, T.; Smoreda, Z.; Ratti, C. Delineating geographical regions with networks of human interactions in an extensive set of countries. *PLoS One* **2013**, *8*, e81707.
6. Golder, S.A.; Macy, M.W. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. *Science* **2011**, *333*, 1878–1881.
7. Golder, S.; Wilkinson, D.; Huberman, B. Rhythms of social interaction: Messaging within a aassive online network. In *Communities and Technologies 2007*; Steinfield, C., Pentland, B., Ackerman, M., Contractor, N., Eds.; Springer: London, UK, 2007; pp. 41–66.
8. Mitrović, M.; Tadić, B. Bloggers behavior and emergent communities in blog space. *Eur. Phys. J. B* **2010**, *73*, 293–301.
9. Tadić, B.; Gligorijević, V.; Mitrović, M.; Suvakov, M. Co-evolutionary mechanisms of emotional bursts in online social dynamics and networks. *Entropy* **2013**, *15*, 5084–5120.
10. Szell, M.; Grauwin, S.; Ratti, C. Contraction of online response to major events. **2013**, arXiv:1308.5190.
11. Bainbridge, W. The scientific research potential of virtual worlds. *Science* **2007**, *317*, 472–476.
12. Ball, P. The physical modelling of human social systems. *Complexus* **2003**, *1*, 190–206.

13. Szell, M.; Thurner, S. Social dynamics in a large-scale online game. *Adv. Complex Syst.* **2012**, *15*, 1250064.

14. Szell, M.; Lambiotte, R.; Thurner, S. Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 13636–13641.

15. Szell, M.; Thurner, S. Measuring social dynamics in a massive multiplayer online game. *Soc. Netw.* **2010**, *32*, 313–329.

16. Klimek, P.; Thurner, S. Triadic closure dynamics drives scaling laws in social multiplex networks. *New J. Phys.* **2013**, *15*, 063008.

17. Szell, M.; Sinatra, R.; Petri, G.; Thurner, S.; Latora, V. Understanding mobility in a social petri dish. *Sci. Rep.* **2012**, *2*, doi:10.1038/srep00457.

18. Thurner, S.; Szell, M.; Sinatra, R. Emergence of good conduct, scaling and zipf laws in human behavioral sequences in an online world. *PLoS One* **2012**, *7*, e29796.

19. Corominas-Murtra, B.; Fuchs, B.; Thurner, S. Detection of the elite structure in a virtual multiplex social system by means of a generalized K-core.**2013**, arXiv:1309.6740.

20. Castellano, C.; Fortunato, S.; Loreto, V. Statistical physics of social dynamics. *Rev. Mod. Phy.* **2009**, *81*, 591–646.

21. Sinatra, R.; Condorelli, D.; Latora, V. Networks of motifs from sequences of symbols. *Phys. Rev. Lett.* **2010**, *105*, 178702.

22. Gallotti, R.; Bazzani, A.; Rambaldi, S. Towards a statistical physics of human mobility. *Int. J. Mod. Phys. C* **2012**, *23*, 1250061.

23. Eagle, N.; Pentland, A. Eigenbehaviors: Identifying structure in routine. *Behav. Ecol. Sociobiol.* **2009**, *63*, 1057–1066.

24. Song, C.; Qu, Z.; Blumm, N.; Barabási, A.L. Limits of predictability in human mobility. *Science* **2010**, *327*, 1018–1021.

25. Gallotti, R.; Bazzani, A.; Esposti, M.D.; Rambaldi, S. Entropic measures of individual mobility patterns. **2013**, arXiv:1305.1836.

26. Krumme, C.; Cebrian, M.; Pentland, A. Patterns of individual shopping behavior. **2010**, arXiv:1008.2556.

27. Gudmundsson, A.; Mohajeri, N. Entropy and order in urban street networks. *Sci. Rep.* **2013**, *3*, 3324.

28. Takaguchi, T.; Nakamura, M.; Sato, N.; Yano, K.; Masuda, N. Predictability of conversation partners. *Phys. Rev. X* **2011**, *1*, 011008.

29. Wang, C.; Huberman, B.A. How random are online social interactions? *Sci. Rep.* **2012**, *2*, 633.

30. Hanel, R.; Thurner, S. A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. *Europhys. Lett.* **2011**, *93*, 20006.

31. Hanel, R.; Thurner, S.; Gell-Mann, M. Generalized entropies and logarithms and their duality relations. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 19151–19154.

32. Cover, T.; Thomas, J. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*, 2nd ed.; Wiley-Interscience: New York, USA, 2006.

33. Feder, M.; Merhav, N. Relations between entropy and error probability. *IEEE Trans. Inf. Theory* **1994**, *40*, 259–266.

34. Fano, R. *Transmission of Information*; MIT Press: Cambridge, USA, 1961.

35. Sinatra, R.; Gómez-Gardeñes, J.; Lambiotte, R.; Nicosia, V.; Latora, V. Maximal-entropy random walks in complex networks with limited information. *Phys. Rev. E* **2011**, *83*, 030103.

36. Chierichetti, F.; Kumar, R.; Raghavan, P.; Sarlós, T. Are Web Users Really Markovian? In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16–20 April 2012; ACM: New York, USA, 2012; pp. 609–618.

37. Labianca, G.; Brass, D.J. Exploring the social ledger: Negative relationships and negative asymmetry in social networks in organizations. *Acad. Manag. Rev.* **2006**, *31*, 596–614.

38. Kivelä, M.; Arenas, A.; Barthelemy, M.; Gleeson, J.P.; Moreno, Y.; Porter, M.A. Multilayer Networks. **2013**, arXiv:1309.7233.

39. Sienkiewicz, J.; Skowron, M.; Paltoglou, G.; Hołyst, J. Entropy-growth-based model of emotionally charged online dialogues. *Adv. Complex Syst.* **2013**, *16*, 1350026.