Drawing without Replacement as a Game Mechanic

Fernando de Mesentier Silva
New York University
Brooklyn, NY
fernandomsilva@nyu.edu

Christoph Salge
University of Hertfordshire
Hatfield, UK
ChristophSalge@gmail.com

Aaron Isaksen
New York University
Brooklyn, NY
aisaksen@appabove.com

Julian Togelius
New York University
Brooklyn, NY
togelius@nyu.edu

Andy Nealen
New York University
Brooklyn, NY
nealen@nyu.edu

ABSTRACT

We introduce several deck of cards and dice models that can be used to represent stochastic outcomes in tabletop games. We analyze these using a toy game introduced as a Micro Combat game. By simulating the outcome of the game with these different models we can analyze them in terms of their salience, disparity, fairness and obfuscation. We expect this analysis to help designers choose the method that best suits their intended experience.

CCS CONCEPTS

• **Mathematics of computing** → *Exploratory data analysis*; *Time series analysis*; • **Applied computing** → **Computer games**;

KEYWORDS

board games, dice, randomness

ACM Reference Format:
Fernando de Mesentier Silva, Christoph Salge, Aaron Isaksen, Julian Togelius, and Andy Nealen. 2018. Drawing without Replacement as a Game Mechanic. In *Foundations of Digital Games 2018 (FDG18), August 7–10, 2018, Malmö, Sweden.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3235765.3235822

1 INTRODUCTION

Stochasticity is a common element in many games. While it is commonly hidden from the player's perception in video games, it is typically evident in tabletop games. Usual mechanisms to realize stochasticity are drawing from a shuffled deck of cards, flipping a coin or spinning a roulette. Arguably, the most common form of stochasticity in tabletop games are dice - so much, that one German word for board game (Würfelspiel) literally translates to dice game.

Dice are used in many different manners in tabletop games. They are an integral part of Roleplaying Games, where feats are accomplished by players when they are successful in rolling above or below a certain number. Monopoly and Clue both determine player movement with the roll of 6 sided dice. War games use them to emulate random chances in combat. Meanwhile, games such as Yahtzee and King of Tokyo have their main mechanics revolve around rolling, freezing, and re-rolling dice.

One defining characteristic of dice is their local independence - each roll of a die has an even distribution, unaffected by previous rolls. That is, rolling a 6 does not change the probability of rolling another 6. However, humans are quick to assume that a random event can't have long runs of the same result, that it needs more alternation to "truly" be random [1]. This is referred to as Gambler's Fallacy. Because of this common fallacy, the results of dice rolls are sometimes perceived as not random enough, as they lack local representativeness [8], meaning that short sequences of dice rolls likely deviate from the distribution expected for larger sequences.

This perceived lack of randomness can lead to mistrust or frustration with the way dice model randomness in games. Designers might therefore provide alternative, non-dice, methods to generate stochasticity. One example for this are Catan: Event Cards [7], an expansion to Settlers of Catan that replaces the dice in the original game with cards. The included deck has 36 cards, each representing one of the possible rolls one could get with two individual 6-sided dice. This mitigates the luck by adding some degree of predictability, and prevents long drought sequences, where certain resources are never rolled. The fact that players observe the cards being draw, and as a consequence know what is still left to be drawn is another feature that is different in relation to dice. The expected outcome of a die is relatively easy to determine: it is an equal probability distribution between all sides of the die, unaffected by anything else. Card draw mechanics, on the other hand, depend on already drawn cards, making it harder to keep track of the exact probability of the next random event. On the other hand, the entropy – or uncertainty – of the next dice roll is constant, while the uncertainty of the next card can vary, again depending on previous cards and the remaining size of the deck. Both of these effects, used properly, can be used obfuscate the future development of the game, allowing a designer to provide a range of real and perceived stochasticity in a game.

1.1 Overview

In this paper we take a systematic look at mechanics referred to as drawing without replacement (as opposed to pure dice, which realize drawing with replacement) which allows game designers to extend their randomness generation mechanism beyond dice-roll

FDG18, August 7–10, 2018, Malmö, Sweden

ACM ISBN 978-1-4503-6571-0/18/08…$15.00
https://doi.org/10.1145/3235765.3235822

equivalent mechanics. After defining some formal measurements we introduce several possible mechanics that produce sequences of random events and analyze them in regards to local entropy, repeated events and local representativeness. One of these mechanics, which we call conveyor belt, is a novice concept we introduce in this work. We will briefly discuss how these mechanisms can be realized, and how they shape the experience of randomness. We also use simulated sequences to provide some quantitative analysis.

In the second experiment we look at a micro combat simulation, and again use simulated games to illustrate how a combination of the different mechanisms can be used to provide different experiences of fairness, disparity and obfuscation. We also demonstrate how these mechanics can be parameterized, which allows us to basically blend them with classical dice mechanics. These parameters then give us a design space in which we can move from independent, simple dice mechanics to highly interdependent drawing mechanics.

2 RELATED WORK / BACKGROUND

Comparing dice with card draw mechanics provides a range of advantages and disadvantage for both mechanics [2]. So our goal here, similar to an earlier analysis on dice mechanics [5], is not to argue for one or the other, but to shed more light on what a particular mechanic can provide. Consequently, we have to ask what role a stochastic element should play in a game?

One role of randomness is to allow for a game to be more balanced, as it allows a weaker player to sometimes overcome a stronger player with the help of randomness [3]. But as player get better at games, they might also prefer random elements that are more subtle, leading to less chaotic swings, and making the outcome of the game rely more on skill. Such a desire could for example explain the move from dice to cards in Settlers of Catan.

Another design challenge is to prevent *inequity aversion* [4, 6], a situation where one player does much better than the other. Ideally, a game should be kept interesting till the end. With the rise in cooperative boardgames this is a particular challenge, as the ideal game should last for a while and then end with a close victory or loss for the players. Yet the game mechanics should be obfuscated enough that it is not immediately evident to the players that the game is balanced against their actions. A draw mechanic can be very helpful here, as it might give the player the illusion that different games have varied developments, as they players might draw better or worse events early on, but are then brought back to an average game performance as the game continues and the remaining events happen. This automatic balancing against earlier good or bad events would be hard to produce with independent dice.

3 FORMALISM AND DEFINITION

The simplest way to express the different mechanics we want to discuss is to imagine we are drawing elements from a multiset, a set that can have duplicate elements. A classical six-sided dice is than represented by the set $D_6 = \{1, 2, 3, 4, 5, 6\}$, where rolling the die will yield one of its entries with equal probability. Classical dice are referred to as drawing with replacement, because in the case of the dice the drawn element will be returned, so the next roll of the dice will be a draw from the same deck.

Drawing without replacement, like the previously mentioned event cards, works by not immediately replacing the element after it has been drawn from the deck. So, if we start with a D_6 and draw a 5, then the next draw will be made from the set $\{1, 2, 3, 4, 6\}$. If we want to generate longer sequences, we also need to define what happens when the deck is empty. For the first example, we assume the deck gets reshuffled when it is empty.

Now, we can immediately see that sequences generated with this drawing mechanic are different from pure dice. For one, the outcome of this dice deck becomes very predictable once it is nearly empty. We can measure this by computing the entropy $H(.)$ at time t of the expected outcome of the event generated from D_t as:

$$H(D_t) = \sum_{r \in D_t} p(r) \log p(r) \tag{1}$$

For the simple sequence without replacement the entropy follows a sawtooth line over time, basically going down to zero at each multiple of 6. Whenever the entropy is low, the next event is easier to predict.

The other effect of using a deck is a higher local representativeness. It is unlikely to generate the same number twice in a row (unless it is the last and first number drawn), and it is impossible to generate the same number three times in a row. With real dice, this is possible at $1/6$ and $1/36$ probability for each point in time. To measure this deviation from the local representativeness we define a measure called salience $S(.)$, that measures how far a given sequence $D_{1..N}$ deviates from the expected values. Given a window size k, which is part of the definition of the measure, we count $c(.,.)$, for each windows of size k in the sequence how probable it is that a certain number appears. So, for a window of size 6, we might count one 1, which leads to a probability of 1/6 for the ones in this sequence. For each number we then calculate the difference between this actual value and the expected value, summing these values for all numbers. This gives us the salience of a given window. By averaging this value for all windows we get an overall measure of how far a sequence deviates from local representativeness.

$$S_k(D_{1..N}) = \frac{\sum_{t=1}^{N-k} \sum_{i=1}^{k} p_D(i) - c(i, D_{t..t+k-1})}{N-k} \tag{2}$$

Finally, we can also measure how likely sequences of n identical events are. For this we basically look at a generated sequence and for each entry check if the next n-1 entries are identical. If we divide this by the length of the sequence we get a probability for how often a sequence of length n appears.

3.1 Analysis of Random Sequences

Now that we defined some measures, we can look at different sequences of random events. We will use this section to introduce several different mechanics to generate sequences with different characteristics. For simplicity, all mechanism will focus on generating numbers from 1 to 6.

We already introduced the first two mechanism, simple dice, and a simple deck of 6 cards, where we refill the deck once it is empty. If we look at the average entropy, we can see that the deck has a low average entropy, and a closer analysis shows that this high predictability always appears at the same time (when the deck is nearly empty).

To counter this, we can raise the entropy by introducing another card into the deck, that, once drawn, triggers a reshuffling of the deck, and the sample is then drawn from the shuffled deck. This reset becomes more likely as the deck becomes more empty. We can also parameterize the deck with a reset card by adding r reset cards. As the amount of reset cards becomes larger and larger, the properties of the deck approximate the properties of a dice, because there is a high chance the deck will be reset before every draw.

Another interesting property is how likely two identical events can happen in succession. This is particularly problematic when we talk about a game where one specific roll of the die has a special effect, and triggering this special effect several times in a row is undesired. The deck of cards guards well against that, but we might want to open the possibility for two, or even more identical events to happen, just with a lower probability. Thus, we may want to blend the simple deck towards being more like a die.

One way to do this is to still use a deck, but duplicate the cards. So, a deck might contain two 1, two 2s, etc. This will make identical sequences more likely. Again, this can be parameterized by increasing the multiplier m of how many identical cards we add. For really large numbers, the behavior approaches that of a simple die.

Finally, we might also want to further reduce the chance for identical results, while still maintaining a low saliency (lower than dice). For this, we can employ a conveyor belt-like mechanism. Instead of putting a card that is drawn back immediately, we put it back with a delay, so the first card drawn is put back after the second card is drawn. If we do this with a card deck we need to reshuffle, but one can imagine this being done with colored stone drawn from a bag. The delay can be realized by using the drawn token also as a marker on the game board, that is then later put back into the bag. This mechanism ensures that no two identical events will happen after each other. Again, we can increase the delay b, and as we do the salience gets even lower. What basically happens is that as the delay reaches the size of the deck it creates a mechanism that just repeats the initial sequence. This means every window of size 6 has exactly one of each number - but at the same time, this will be then a fully predictable sequence.

There are two things to note. One, all these mechanics can be combined, so for example it is possible to have a deck with $m = 3$ duplicates of all cards, $r = 6$ reshuffle cards, and also a conveyor belt mechanic with a delay of $b = 2$. This allows for a parameterization of $D_{m=3,r=6,b=2}$ encoding different mixed mechanics. Actual dice do not have to be formalized separately, as they can be expressed as $D_{1,0,0}$ as a delay of 0 leads to immediate replacement.

3.2 Discussion of Sequence Results

If we look at table 1 we can see how likely sequences of a given length are to appear with the different methods. Regular dice produce the most sequences, there probability decays, as expected with $1/6, 1/36, \ldots$. Note that this, as indicated by literature, is actually more than player would naively expect. Using a simple deck instead removes all sequences longer than 2, and only allows those at the point where the deck gets refilled, which might appear a bit to limiting, if they designer want the rare occurrence or threat of 3 element sequences. Both reshuffle and duplicates allow here for a blending of the two approaches. The deck with duplicates still introduces a hard limit of 4, and makes shorter sequences more likely than the deck, but still less likely than the dice. A deck with a reshuffle card allows, in theory, for sequences of unlimited length, but anything beyond 4 seems to be so improbable that it will likely never be seen. The conveyor belt is even more extreme in this metric, at is completely removes all sequences.

3.3 Salience

Salience indicates how much a local window of results deviate from the overall expected distribution. I gives an indication of how odd a given sequence might appear to players. Table 2 shows the value of salience of the different mechanics over windows of different sizes. Dice, as previously discussed, deviate quite a lot from the law of small numbers, indicated by the overall high values. Decks all have generally lower values, as the appearance of a given number makes it less likely to appear again soon. The conveyor belt with different length give us the option to create a range of different saliency values, from nearly like dice for a delay of 2, to very low saliency sequences with length 5. It should be noted thought, that those sequences are highly predictable, as they basically repeat the same initial sequences over and over again. Also, as the window grows larger we see that all sequences get closer to their expected values, but different mechanisms drop of at different rates.

4 METHOD AND METRICS

To perform our evaluations, we first need to define the game being used as benchmark and the metrics used for evaluation. The game was purposely selected to be simple, as to make more evident the differences in the mechanics.

4.1 Micro Combat Game

In order to perform experiments, we created a toy game designed to be an extreme simplification of combat in a tabletop roleplaying game.

The game has 2-players fighting to be the first to reduce their opponents health points (HP) to zero. Both players start with the same number of HP. Gameplay consists of players alternating turns doing damage to their opponent.

There are no actions for players to choose from. On a players turn they use a stochastic element to decide the amount of damage they are inflicting. If this was used a a game mechanics, then the relevant decision of the player would be around this event, as in does the player want to enter this combat, or does the player want to use additional resources to change the odds.

Through our experiments we will be evaluating the design impact of modifying the mechanism used to simulate the stochastic decisions. We will also evaluate the results under different amounts of starting HP. The mechanic of the game results in imbalance between players. The first player has the upper hand from always being 1 turn ahead of their opponent. This is another point of observation to highlight as we alter the feature's design.

4.2 Metrics

To analyze the impact the changes have on the design, we came up with 2 metrics of evaluation.

	Length of Sequence					
Method	**2**	**3**	**4**	**5**	**6**	**7**
Dice	0.1677	0.0283	0.0048	0.0008	0.0002	0.0000
Deck	0.0283	0.0000	0.0000	0.0000	0.0000	0.0000
Deck with Reshuffle Card	0.0977	0.0026	0.0001	0.0000	0.0000	0.0000
Deck with Duplicates	0.0463	0.0013	0.0000	0.0000	0.0000	0.0000
Conveyor-belt with delay of 2	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Table 1: Table showcasing the frequency of sequences of same result drawn in a row. Method indicates which mechanic is being used. The length of sequence label indicates how many repetitions in a row are necessary to be considered a sequence. The values in the columns indicate the probability of observing a sequence of such length, counting for all numbers.

	Window Size			
Method	**6**	**12**	**24**	**384**
Dice	0.403	0.271	0.187	0.046
Deck	0.194	0.089	0.042	0.003
Deck with Reshuffle Card	0.264	0.165	0.110	0.026
Deck with Duplicates	0.299	0.137	0.065	0.004
Conveyor-belt delay of 2	0.327	0.220	0.152	0.038
Conveyor-belt delay of 5	0.099	0.068	0.048	0.012

Table 2: Representation of the Salience of the different methods for various window sizes. Rows represent the different mechanics, while columns refer to the size of each window. Values are calculated according to equation 2.

Disparity: Designed to measure how close was the losing player to actually win. It is modeled after the common discussion players have over the outcome of the game. It is usual for the defeated player to reach an analysis of how many turns away they were from winning. Our disparity metric is then measured as number of turns. By simulating extra random outcomes, we can determine how many more turns, in average, it would take for them to win the game. Matches with close scores have low disparity values.

Fairness: Evaluates how likely a player is to win. As mentioned before, the game we are using as benchmark favors the first player. This is a common problem in tabletop design for turn based games. We are proposing that tuning the starting HP and stochastic simulation methods can have an impact on the degree of unbalance towards the starting player. We evaluate fairness by simulating a large number of games and measuring the win ratio of the starting player. The closer the metric is to 50%, more balanced the game is.

Aside from these metrics, we discuss a subjective manner of evaluation. Through it, we intend to relate the changes in stochasticity model to the predictability of the result of a match. We define this evaluation as:

Obfuscation: Represents the author's subjective evaluation of the human player's ability to perceive the actual distribution of the stochastic model. Arguably, we can indicate it as a factor of how predictable the method is versus how much memory is necessary to accurately calculate the outcome. The more obfuscated it a method is, the harder it is for humans to perceive the actual distribution.

5 EXPERIMENTS

In this section we will discuss the results of our experiments and analyze the relation between metrics and methods. To get results we run 100,000 simulations for each instance (each initial HP being evaluated) of the game. Since there are no decisions to be made on part of the players, it is not necessary to have agents to play the game.

All methods in the experiments come from the mechanics introduced in sections 3 and 4. During our results discussion, each player is generating results from their independent component. Dice refers to rolling a 6 sided die. Deck means drawing from a deck of 6 cards, each with a unique value from 1 to 6, being reshuffled after depleting. Deck with reshuffle is equivalent to Deck, but with one extra card called reshuffle. When a player draws such, the whole deck is reshuffled and the same player proceeds to drawing a new card. Deck with Duplicates work as the regular Deck, but starts with 2 copies of each card (making it a 12 card deck). Conveyor belt delay 2 means a deck of 6 cards, where in the beginning of every turn the card discarded 2 turns ago gets reshuffled into the deck.

5.1 Disparity

Our first experiment is to analyze disparity. As presented before, disparity measures how many more of their turns were necessary for the losing player to reduce the opponents HP to 0. Results for this experiment are shown in Figure 1.

When analyzing the results for the Deck mechanic we can notice a repeating pattern over a constant Initial HP interval. This is expected as over the course of 6 turns players will always cause a total damage of 21 to their opponents, which is equivalent to the sum of the value of all cards in the deck. Analogously, the Deck with Duplicates repeats the same pattern with twice the amplitude for the pattern, as expected.

When rolling a 6 sided die, increasing the initial HP has the variance in outcomes push the disparity to high numbers, meaning wins with comfortable margins. In terms of design this might not be desirable, as it creates a snowball effect, where the player that has to close the gap encounters little motivation to do so.

With similar curve shape, Deck with Reshuffle and the Conveyor belt of length 2, seem to be more reliable mechanics to maintain the

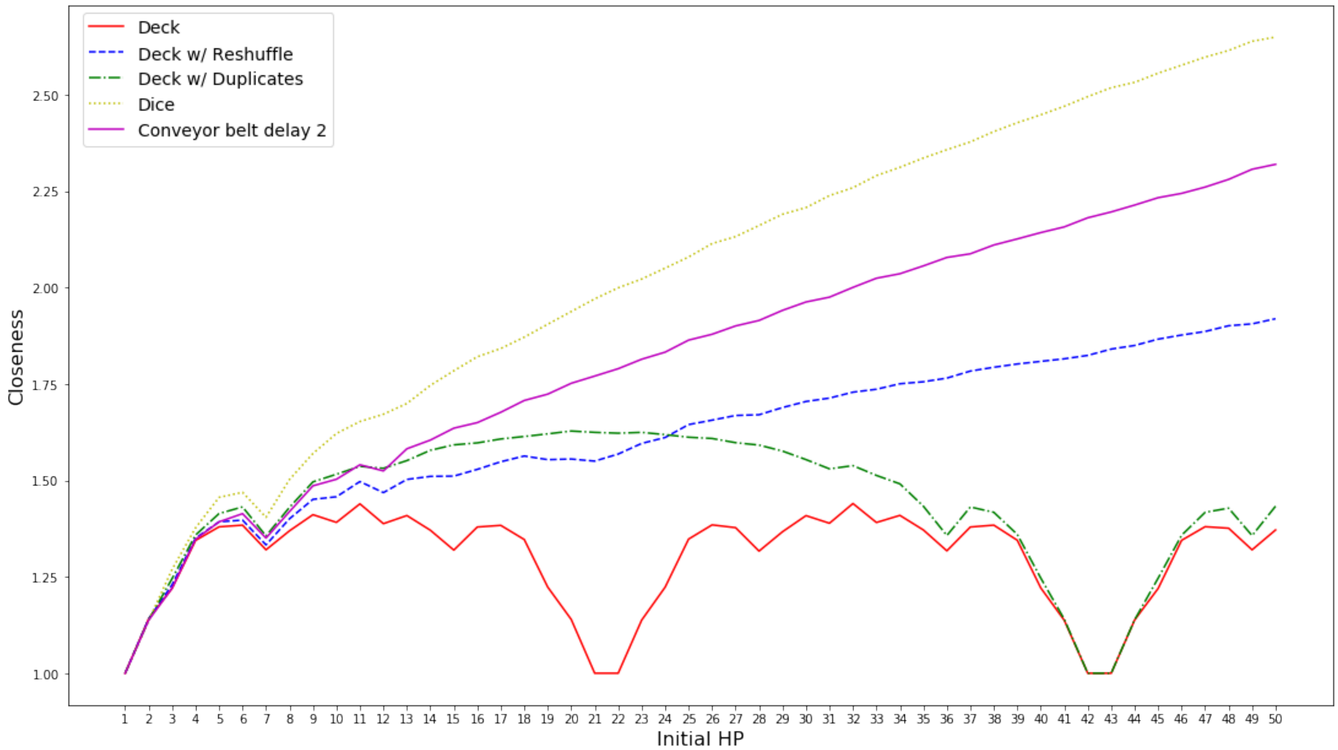

Figure 1: shows the results of the simulations in terms of disparity. The Y-axis represents the initial HP players had for an instance, while the Y-axis maps the disparity of the results.

players closer in score when compared to dice. Despite showing very similar behavior, the discarded values help reduce the variance present on the Dice, giving the change of the player which is trailing to draw higher value cards, while the player that is currently ahead has a higher chance to draw lower numbers.

5.2 Fairness

Fairness, as mentioned before, is calculated as the win ration of the first player to act. The closer the game is to 0.5, the more fair it is. Since the game is naturally unbalanced, we are interested in investigate how the different mechanics impact the advantage the starting player has.

The first thing to notice is how the variance works in favor of Fairness. The methods with the highest variance in results are all more successful at making the game more balanced. That said, a significant amount of initial HP is necessary to provide enough turns for the variance to benefit the second player to act. Despite not being the target of the toy game discussed in this work, it is worth noting that prolongating a game in this fashion, could have a negative impact of the player's enjoyment of the game.

It is also clear, as we compare Figures 1 and 2 that, the methods that make the game more fair, are also the ones that have, in average, matches that end with the winner in a more comfortable lead. Designers can use this information to choose their stochastic mechanic in terms of the desired trade-off between both. If opting for methods that can provide both a more balanced play with tighter matches, the Conveyor belt and Deck with Reshuffle seem to be the most interesting candidates.

6 OBFUSCATION

Looking at disparity and fairness, we can see that they are anti-correlated to some extent, giving us games with very similar properties for different hit points. Take, for example, the deck games with 3 and 19 initial hit points. While they do have the same values, they provide a different experience. One is a very obvious advantage for player one, all player one needs to do is draw 3 or higher, and win on turn on. The game with 19 hit points actually works very similar - basically all player one has to do is to not draw a 3 or higher as the last card in the deck. There is a symmetry here induced by the slowly emptying deck, but this is much harder to see for a casual player. Being asked if one wanted to rather play a deck game as the second player with 3 or 20 points of health would not be a trivial question. This is what we mean by obfuscation - the game with more hit points is more obfuscated, as the player can not as easily model the outcome distribution. This might be useful for a game designer, that might want to create a game experience that is both close and massively biased towards the player without obviously appearing so.

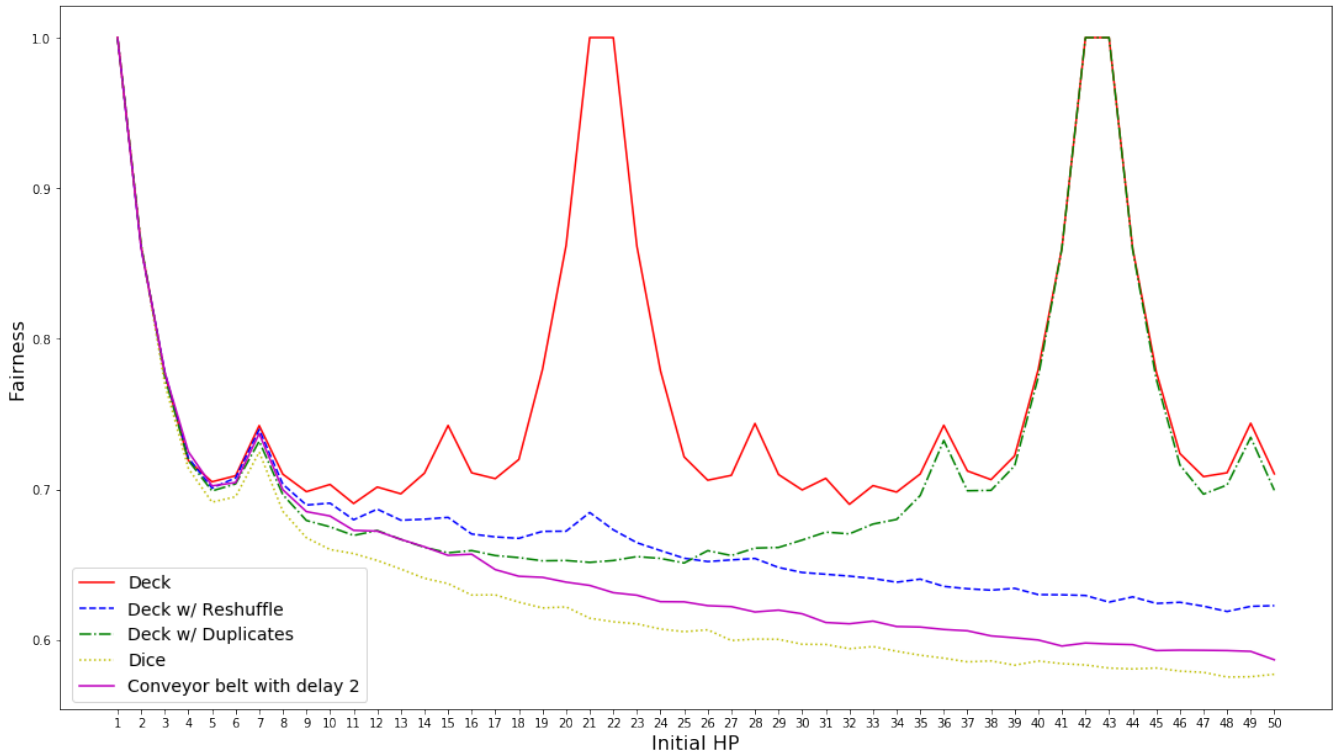

Figure 2: shows the result of Fairness over different initial HP values. The X-axis represents the different initial HP that were simulated. The Y-axis measures the fairness of such game in the specific mechanic.

7 DISCUSSION AND CONCLUSION

We presented methods to represent stochasticity in tabletop games, analyzing them in terms of Salience, Fairness, Disparity and Obfuscation. With this analysis we hope to help designers choose the mechanic that suits the intended design experience.

While Dice are the highest on salience, it is the mechanic that allow for the most fair of games, under the toy game we present. Although, due to its variance, the average matches are the most distant in terms of disparity. Counter intuitively, the game we discussed had an anti correlation between disparity and bias, so the games had the tightest outcome in terms of time also where the most unfair.

Decks, with or without duplicates, have more predictable outcomes, that allow players to plan for, but as a consequence highlight the intrinsic imbalance of the game. Meanwhile, Deck with Reshuffle and the Conveyor belt with delay 2 are the mechanics with the least trade-off between close matches in average and bias towards the starting player, but as a consequence are the most obfuscated.

8 FUTURE WORK

As we analyzed the impact of these different methods on a toy game, we would like to expand it to a more robust game. We are also interested in analyzing human reception of such mechanics by conducting a user study, which would help verify our assumptions about obfuscation. Furthermore, it would also be interesting to automatically explore the parameterized space of dice mechanics, performing and expressive range analysis on what different stochastic experience metrics we can reach.

ACKNOWLEDGMENTS

Authors thank the support of CAPES, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brazil. CS is funded by the EU Horizon 2020 programme under the Marie Sklodowska-Curie grant 705643.

REFERENCES

[1] Maya Bar-Hillel and Willem A Wagenaar. 1991. The perception of randomness. *Advances in applied mathematics* 12, 4 (1991), 428–454.

[2] Scott Caputo. 2015. Mechanics Face Off: Dice vs. Cards. (apr 2015). http://www.leagueofgamemakers.com/mechanics-face-off-dice-vs-cards/

[3] George Skaff Elias, Richard Garfield, and K Robert Gutschera. 2012. *Characteristics of games.* MIT Press.

[4] Ernst Fehr and Klaus M Schmidt. 1999. A theory of fairness, competition, and cooperation. *The quarterly journal of economics* 114, 3 (1999), 817–868.

[5] Aaron Isaksen, Christoffer Holmgård, Julian Togelius, and Andy Nealen. 2016. Characterising Score Distributions in Dice Games. *Game and Puzzle Design 2.1* (2016).

[6] Aaron Isaksen, Mehmet Ismail, Steven J Brams, and Andy Nealen. Catch-up: A game in which the lead alternates. (????).

[7] Klaus Teuber. 2005. Catan: Event Cards. *https://boardgamegeek.com/boardgameexpansion/20038/catan-event-cards Accessed on 05-30-2018* (2005).

[8] Amos Tversky and Daniel Kahneman. 1971. Belief in the law of small numbers. *Psychological bulletin* 76, 2 (1971), 105.