A Statistical Analysis of The End

Anton Strezhnev

May 31, 2019

Contents

1 Introduction

In May of 2017 Swim Pony Performing Arts premiered the first playthrough of its 28-day interactive theatre project, *The End*. Unfolding over a series of text message conversations between the participants and a mysterious entity called "The End," the experience is a broad meditation on mortality, death and meaning of life. Each day, the participants or "journeyers" are guided by "The End" to select a card from a deck which then tasks them with a particular activity exploring mortality.

The End is unconventional not only in its structure, which draws heavily on elements of alternate reality game (ARG) design, but also in the amount of information that it generates regarding participant experience. Because interactions between participants and actors are conducted via text-message, Swim Pony has available to it a rich dataset of the entire output of the game. Few theatrical projects can claim to possess real-time information on how its audience is experiencing the performance.

This data lends itself to asking a variety of interesting questions. One of the core motivations behind the game is based on research in psychology regarding the positive impact of contemplating mortality on happiness and emotional wellbeing.[1] To what extent is this actually occurring among players of the game? Additionally, the data allows Swim Pony to evaluate *The End* with the aim of improving game design in future iterations of the project. When are players most active? How often do they play? When are players most likely to drop off? What cards or experiences are associated with greater or lesser activity? Questions regarding variation in player experience are also highly relevant. The aim of the project is to be as accessible as possible to as diverse a pool of participants as possible, and it is important to learn about any possible heterogeneity in player experience. Considering the emotionally taxing theme of mortality, it is particularly useful to assess whether players with histories of depression or other mental health interact differently with the actors. Finally, Swim Pony wants to know how players "grew" in their interactions with "The End" - did the experience achieve its intended goal of opening players up to think more comprehensively about death and mortality?

This report provides an initial analysis of the data from the May 2017 performance. A total of 57 players participated from May 1 to May 28. Of these players, 56 had accessible text chat logs. Of these 56 sets of

[1] See for example: DeWall, C. Nathan, and Roy F. Baumeister. "From terror to joy: Automatic tuning to positive affective information following mortality salience." *Psychological Science* 18, no. 11 (2007): 984-990., Kashdan, Todd B., C. Nathan DeWall, David R. Schurtz, Timothy Deckman, Emily LB Lykins, Daniel R. Evans, Jessica McKenzie, Suzanne C. Segerstrom, Matthew T. Gailliot, and Kirk Warren Brown. "More than words: Contemplating death enhances positive emotional word use." *Personality and Individual Differences* 71 (2014): 171-175.

logs, 52 were manually coded by Swim Pony. Coders assigned labels to individual messages from participants that exhibited engagement with game mechanics, displays of vulnerability, personal reflection, and efforts at taking agency and action. As part of the game on-boarding players took surveys that asked questions about their mental health, time availability, readiness for a variety of tasks, etc. Around two thirds of players also completed a post-game survey that included a repeat of the mental health questions. The remainder of the report is structured as follows: Section 2 evaluates the factors explaining player engagement and activity in *The End*, Section 3 builds on this to evaluate patterns of drop-out and player disengagement, Section 4 considers the relationship between player characteristics and activity levels, Section 5 then evaluates the hand-coded message data to see whether player characteristics explain different types of player engagement beyond general activity.

2 Summary of participants

This section provides an overview of the survey responses provided by the players during the intake and post-game debriefing periods. Of the 57 participants in the game, we have 55 who answered demographic and background questions prior to the start of the game.

2.1 Demographic characteristics

We first summarize the general age and gender distribution of respondents who participated in the game.[2] Overall, players in this initial round tended to be women, primarily in the 30-49 age group. A fair number of younger players also participated, but there were very few who were older than 50.

Table 1: Player counts by age and preferred gender pronoun

	18-29	30-49	50-64	65+
she/her	12	21	5	1
he/his	4	4	2	1
other	1	3	0	0

Players also answered questions regarding their mental health and experience with loss (Figure 1).

[2]We omit race from this particular summary since in this particular iteration of the game, the vast majority of respondents who volunteered were white.

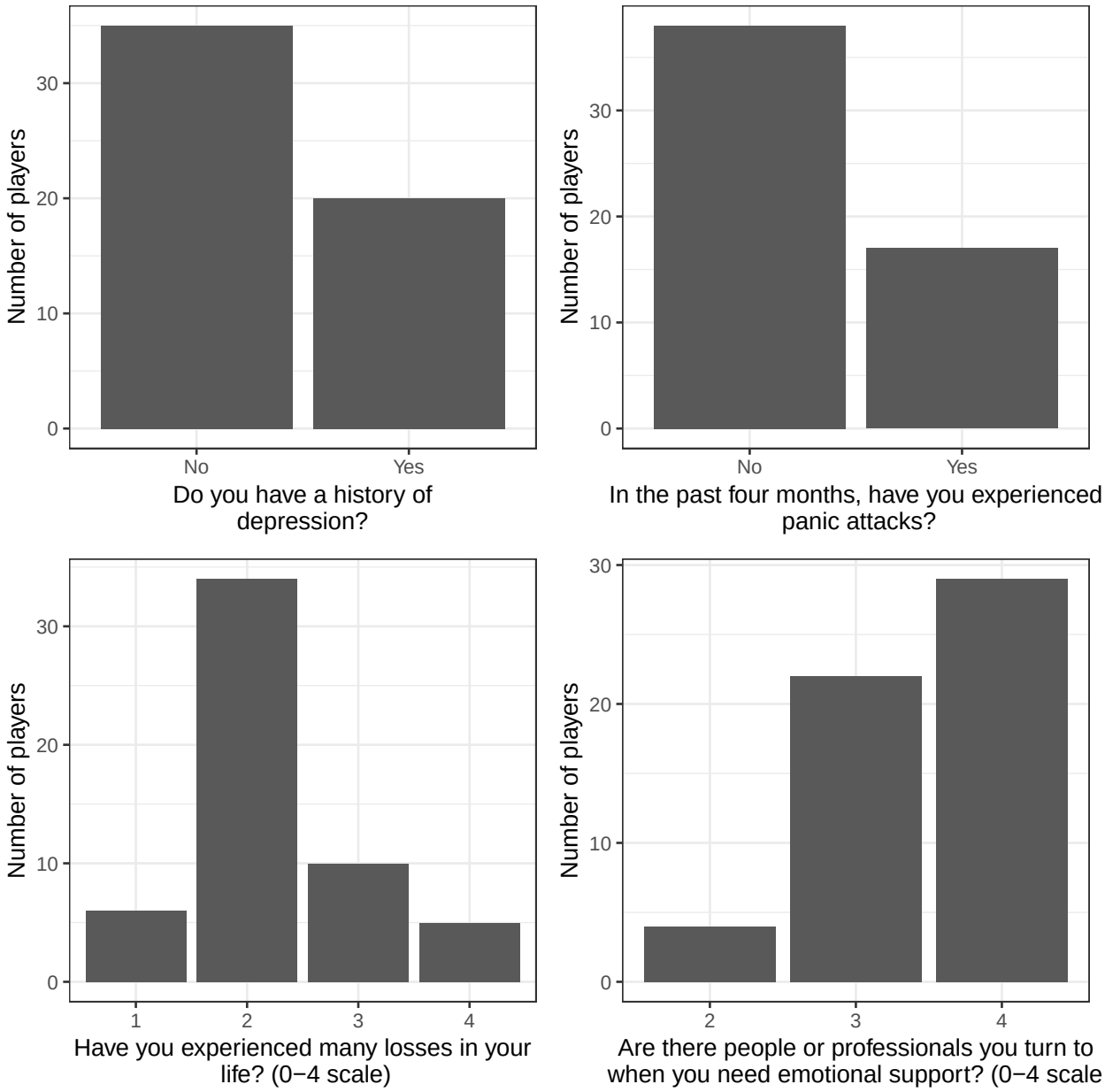

Figure 1: Players' responses to mental health questions

2.2 Flourishing metrics

In addition to the basic demogrpahic information, players also responded to six questions designed to measure "flourishing" - a concept in positive psychology developed by Martin Seligman.[3] Seligman conceptualizes human "flourishing" along five dimensions: positive emotion, engagement, relationships, meaning and accomplishment (PERMA). Respondents answered questions related to a subset of these dimensions taken from a broader 23 question PERMA-Profiler battery developed by Butler and Kern.[4] The game designers selected the following

[3]Seligman, Martin E.P. Flourish: A visionary new understanding of happiness and well-being. Simon and Schuster, 2012.

[4]Butler, Julie, and Margaret L. Kern. "The PERMA-Profiler: A brief multidimensional measure of flourishing." International Journal of Wellbeing 6, no. 3 (2016).

six questions, emphasizing in particular measures related to self-reported "meaning" in life:

Table 2: Flourishing questions

Dimension	Question
Engagement	How often do you lose track of time when doing something you enjoy?
Relationships	To what extent do you receive help and support from others when you need it?
Meaning	To what extent do you generally feel you have a sense of direction in your life?
Meaning	In general, to what extent do you lead a purposeful and meaningful life?
Negative Emotion	In general, how often do you feel anxious?
General Happiness	Taking all things together, how happy would you say you are?

Each of these questions was measured on a 0 to 10 scale with 0 anchoring the lowest possible value (e.g. "not at all" or "terrible") and 10 the highest ("completely" or "excellent"). Thirty-nine of the respondents answered these questions both prior to the game and at the conclusion of the experience. We compare the average responses of these respondents to assess whether the players' self-reported emotions changed over the course of May. Table 3 presents the average pre- and post- responses for those players who answered both survey waves along with the results of a paired t-test for the null of zero difference between pre- and post-game waves.

The largest two changes in the flourishing metrics appeared in the "sense of direction" and "anxiety" questions. Respondents post-*The End* exhibited lower levels of self-reported anxiety and a greater sense of direction in their lives. These differences are distinguishable from statistical noise at a $p < .05$ rejection threshold. We also find greater willingness of respondents to say they receive help and support when they need it. Conversely, we find no meaningful change in the extent to which respondents believe they "lose track of time" when doing things they enjoy. We do also find some shift in the "purposeful and meaningful life" measure, though this may be attributable to statistical noise. Finally, we see about a .5 point increase in self-reported happiness, which corresponds to roughly a one-third standard deviation increase above the pre-game average.

Table 3: Player flourishing metrics before and after 'The End'

Question	Avg. (Pre)	Avg. (Post)	Change	t-Test p-value	# Respondents
Lose track of time	7.3590	7.3846	0.0256	0.9431	39
Receive help and support	7.1053	7.6579	0.5526	0.0160	38
Sense of direction	6.5000	7.3421	0.8421	0.0001	38
Purposeful life	6.8947	7.2368	0.3421	0.1133	38
Feel anxious	6.1622	5.4054	-0.7568	0.0112	37
Happiness	7.0263	7.5000	0.4737	0.0371	38

It would be improper to assign a causal interpretation to these results considering the absence of a proper control group that did not experience the game coupled with the multitude of factors that could have affected players' emotional state during the month such that respondents pre-May responses cannot serve as proper controls. However, the results are suggestive that there may be some positive impact to the overall experience and are worth potentially evaluating further in a truly randomized trial.

3 Participant engagement and activity

This section considers when and how often players are engaging with the game over the course of the experience. It evaluates typical activity levels throughout any given day, assessing whether there are notable differences as the game progresses/across background characteristics of the players. It then considers activity

throughout the course of the game, assessing on a week-to-week level whether certain days of the week are unpopular and whether this shifts as the game progresses. Finally, it develops card-level measures of popularity, evaluating whether certain cards receive significantly less attention than others throughout the game.

3.1 Activity patterns during the day

During what time of the day are players playing the game? On each active day during the course of the game, the messaging system would contact players at 7:30 AM. Messaging would close at 11:30 PM - between that period, players could interact with actors who would respond as the character "The End" to the participants. Figure 2 plots a histogram of the number of player interactions (messages sent by players) by time for all days in the sample (half-hour bin widths). In general, we see a typical work-day trend. Player activity decreases slightly over the course of the day (from 7:30AM) and increases steadily after noon, with a spike in activity at 4pm. We see another upward trend in activity starting at 9pm and going until the close of the game. The amount of activity around this 9:30pm-11:30pm period is about 3 times the amount we observe during the day, suggesting a need for more actor staffing during these high-volume time periods.

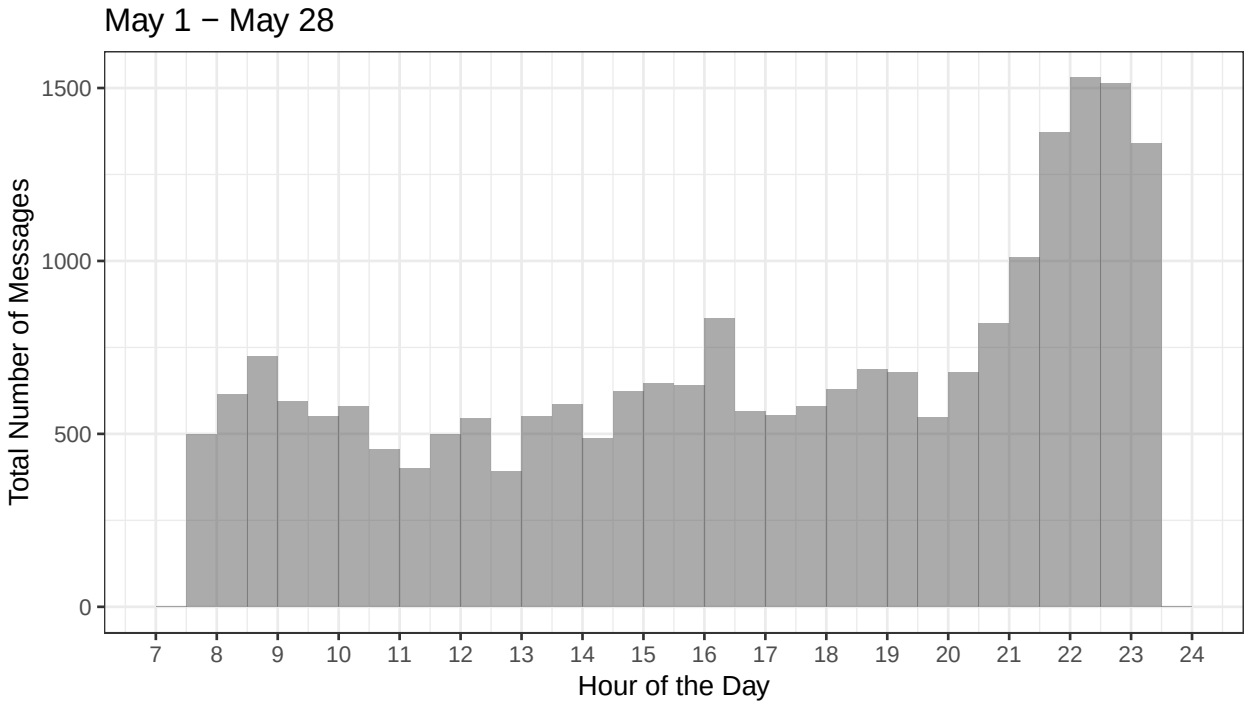

Figure 2: Histogram of daily play frequencies (56 players, half-hour bins)

We find that this distribution in play frequency fluctuates somewhat from week to week. Figure 3 plots the same activity histograms on a week-by-week basis. One important difference between week 1 and the later weeks is the reduction in morning activity. Week 1 also does not exhibit the same 4pm spike in activity. Finally, Week 4 sees a more flat activity pattern during the day relative to Weeks 1-3.

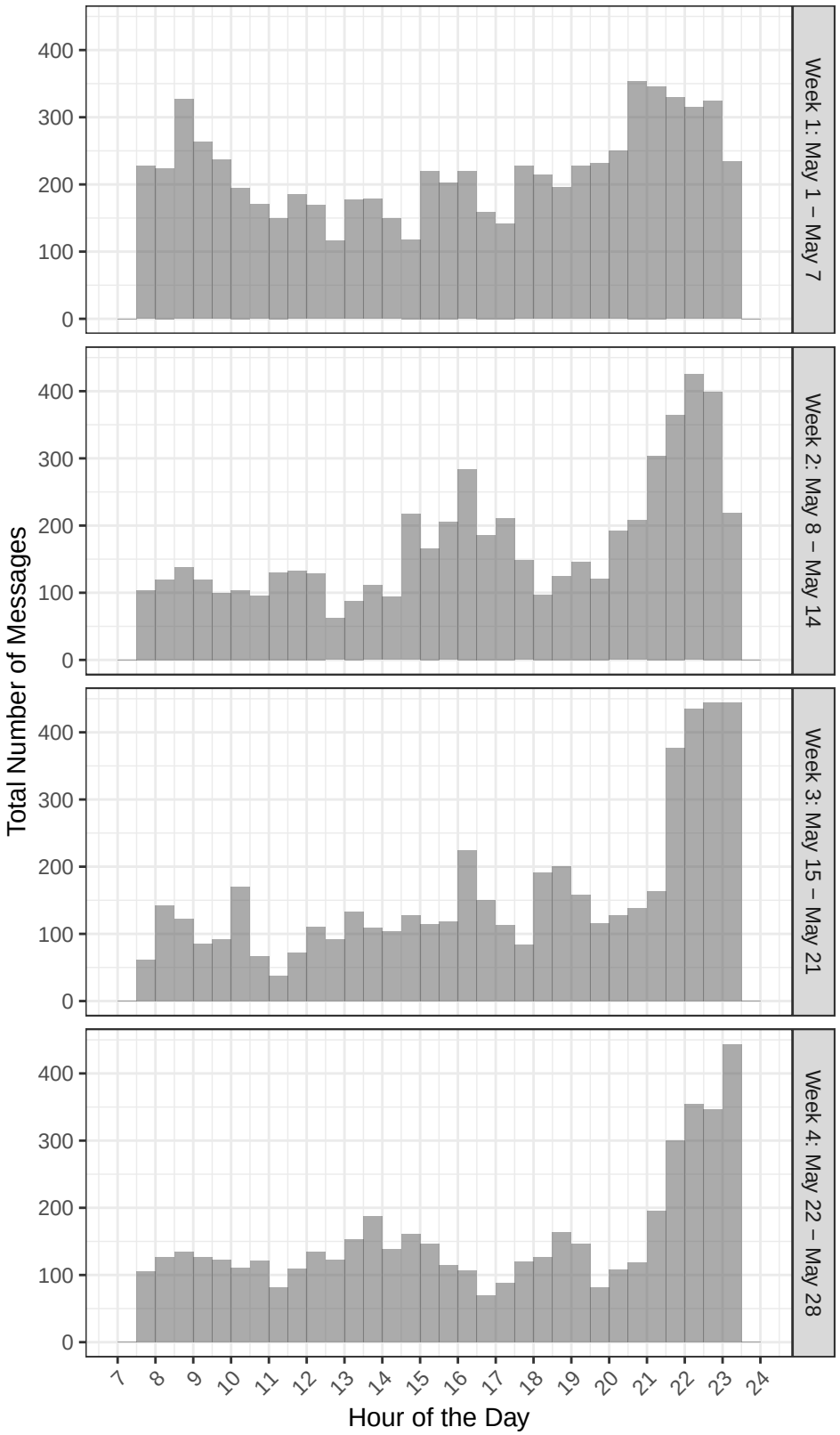

Figure 3: Histogram of daily play frequencies by week (56 players, half-hour bins)

3.2 Activity patterns across days

While there is considerable fluctuation in activity time within individual days, we note that there is very little day-to-day variation in overall message amounts across different week days. Figure 4 plots the total number of interactions for each of the seven days of the week. Mondays exhibit the greatest number of interactions (with a slight dip on Tuesdays/Wednesdays and Sundays), suggesting that work-week cycles are affecting the extent to which players are able to engage in the game. The Sunday-Monday drop-off though suggests that players are using the last day of the week to take a break from the game rather than to engage more with it.

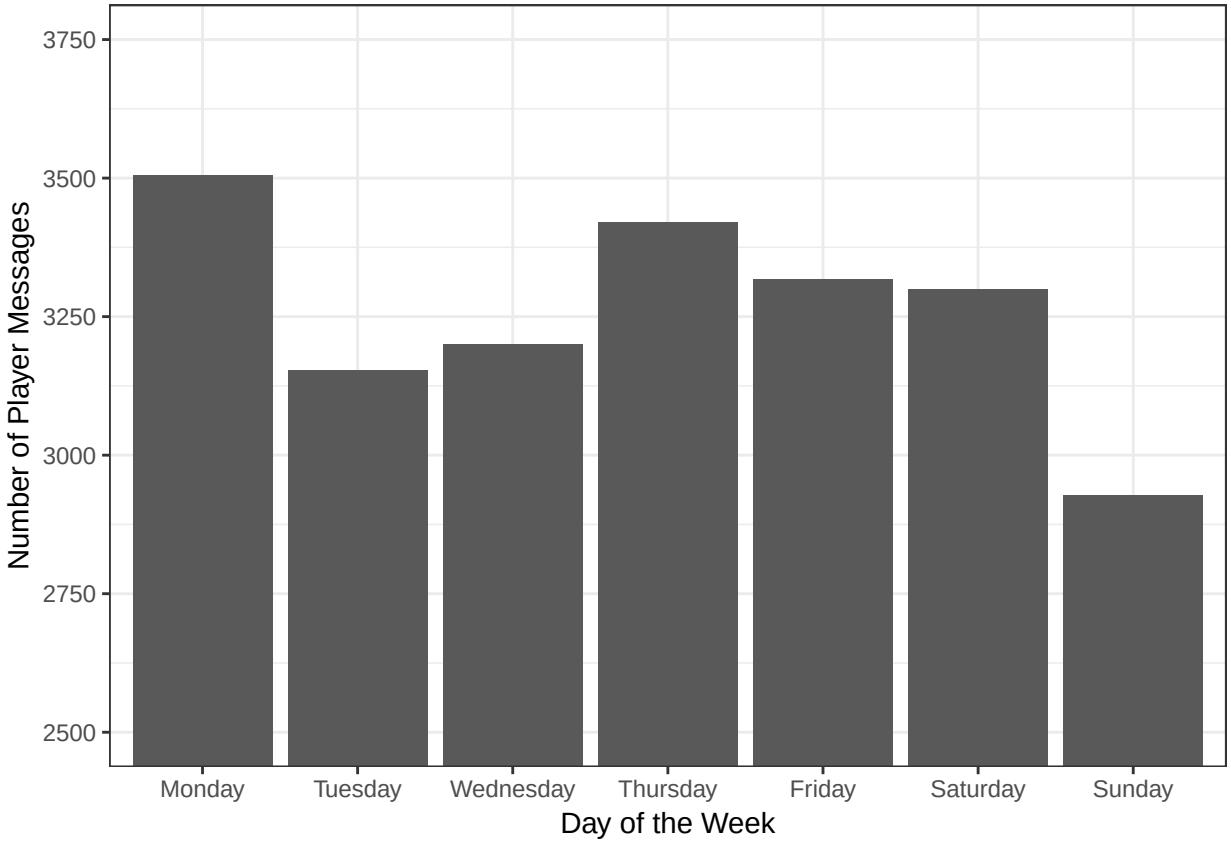

Figure 4: Count of total number of messages by weekday (56 players)

To finally visualize the total magnitude of interactions over the course of the game, we can generate heat-map of interaction frequencies by calendar day. Figure 5 displays the intensity of interaction frequencies for each day. Days in red have an above-average level of interaction while those in blue have a below-average level. As expected, the magnitude of interaction declines over time, but there are a few interesting outliers. During the first week, and particularly the first day, we see an above-average number of interactions. This is likely due to the increased structure built in to the first few days of cards, requiring more back-and-forth between players and *The End* to establish and clarify the rules of the game. Otherwise, the general downward trend is punctuated by a few days of more intensive activity - May 11th being one notable example of an above-average day. Some outliers towards the end are also of interest, and likely reflective of the (Card 38). However, the decline in overall number of messages towards the end of the game may not necessarily be due to fatigue or loss of interest. As players begin to master the reflection and engagement process, they may need less conversation to get to the core of the

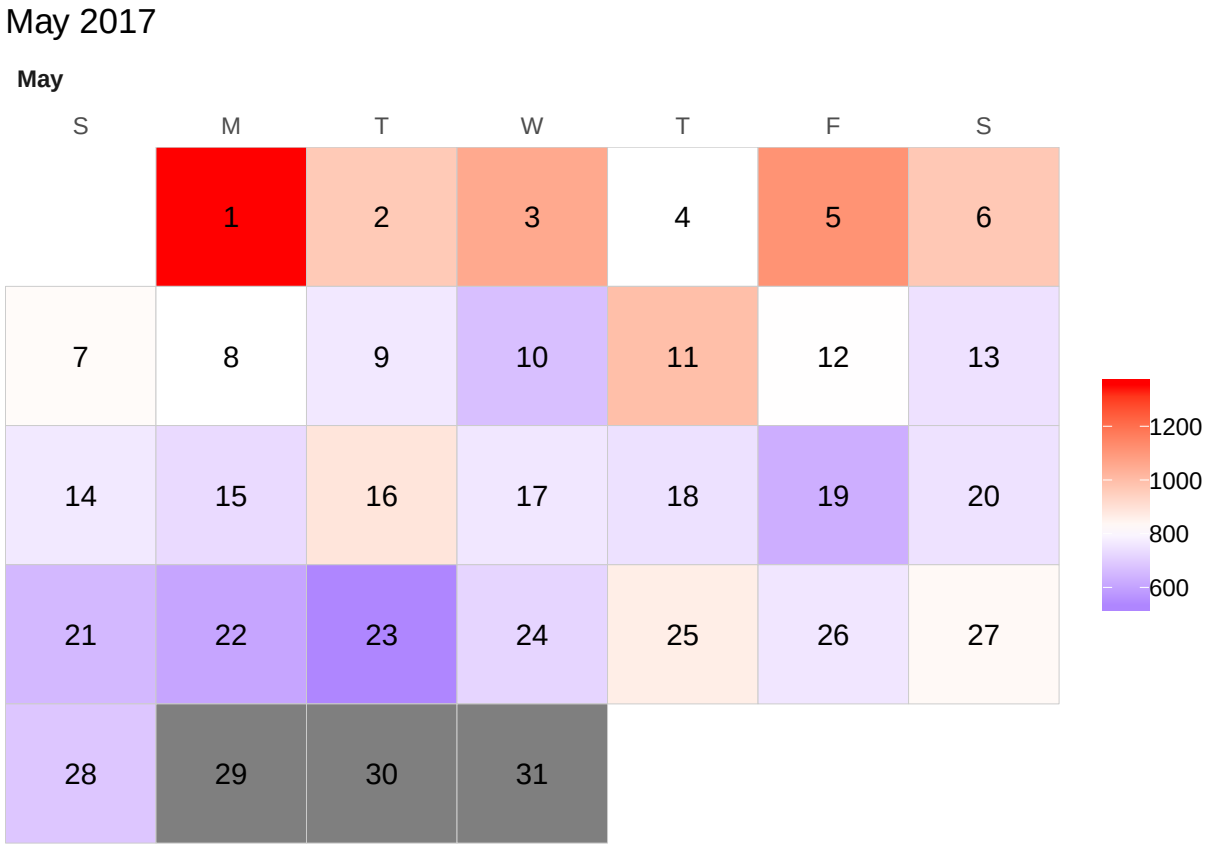

Figure 5: Daily number of player messages

3.3 Activity patterns across cards

The final dimension across which we can visualize activity volume is to consider which cards received more or less attention. On each day, a player could play one or more cards (typically one). For certain days, the player would be guided to a specific card by "The End" (such as on the first day where all players played cards 1 and 2). In other cases the player would be allowed a choice, with some suggestions made by the actors. The game was segmented into three "decks" of a total of 52 cards. Cards 1-13. constituted the first deck. Cards 14-38 the second deck and Cards 39-52 the final third deck.

Figure 6 displays a heatmap with the percentage of players who played each of the 52 cards. In each deck of cards, there is a sequence that is played by everyone (e.g. during the transition between decks 1 and 2, all players play the sequence 12-14). We also see players in the first deck playing most of the cards as there are fewer options available to the players. In Deck 2, players began to have more choices in terms of what cards they could play and we see more dispersion in play frequency. Within a given deck, we find some interesting variation in which cards are rarely selected and which are more likely to be chosen. Notably, when players were given their first choice of card (between card 5 or card 6), Card 6 is substantially more likely to be chosen. Card 33 in particular is very rarely played, which is to be expected given that it is designed only to be played at a particular date and time. The completion card, card 52, was played by about three quarters of players. Interestingly, we find some suggestions that the actors who interact with the players may have guided players to certain cards. We see more than 50% of players playing cards 20, 28 and 28 in Deck 2, which it turns out were created early in the process and were very familiar to the actors. In the future, it may be important to develop "suggested orders" for play and communicating these to the actors to ensure that they are familiar with all of the cards. This may help to avoid implicit bias on the part of the actors towards certain "familiar" cards and away from those developed later in the process.

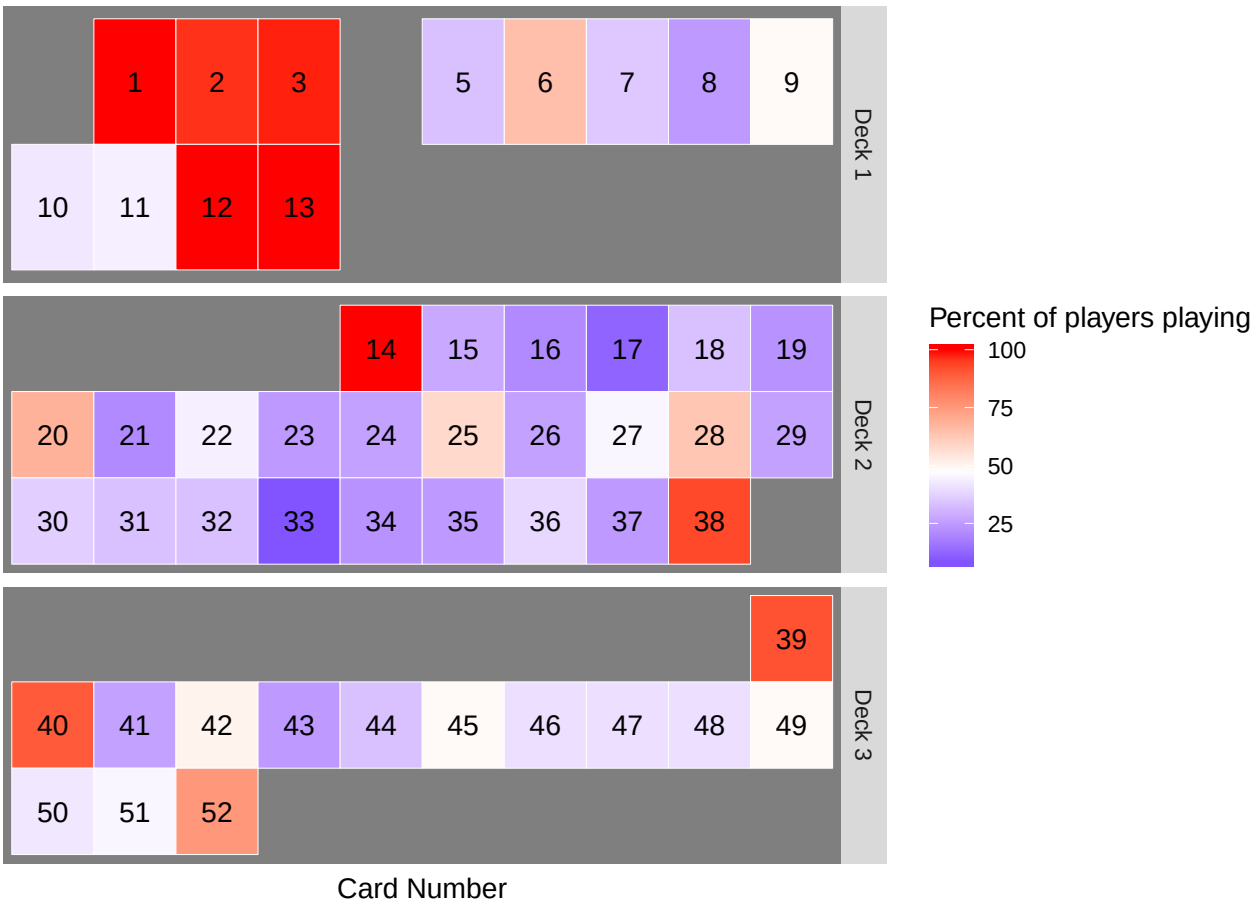

Figure 6: Percentage of players who played each card

In subsequent sections we will address the topic of card "type" and its association with measures of activity and interaction in greater detail. However, for this section we will consider whether players were more likely to be drawn to and select cards with certain types of demands. The cards can be grouped into four types of interactions: writing/research prompts, live conversations, meditative reflection and embodied experience.

Table 4: Sample card tasks

Card Type	Sample task
Writing	Write an imaginary obituary for yourself.
Live Conversation	Have a conversation with a friend or family member about end of life medical wishes.
Meditative Reflection	Think about what the last song you'd like to listen to would be.
Embodied Experience	Roleplay an end of life diagnosis.

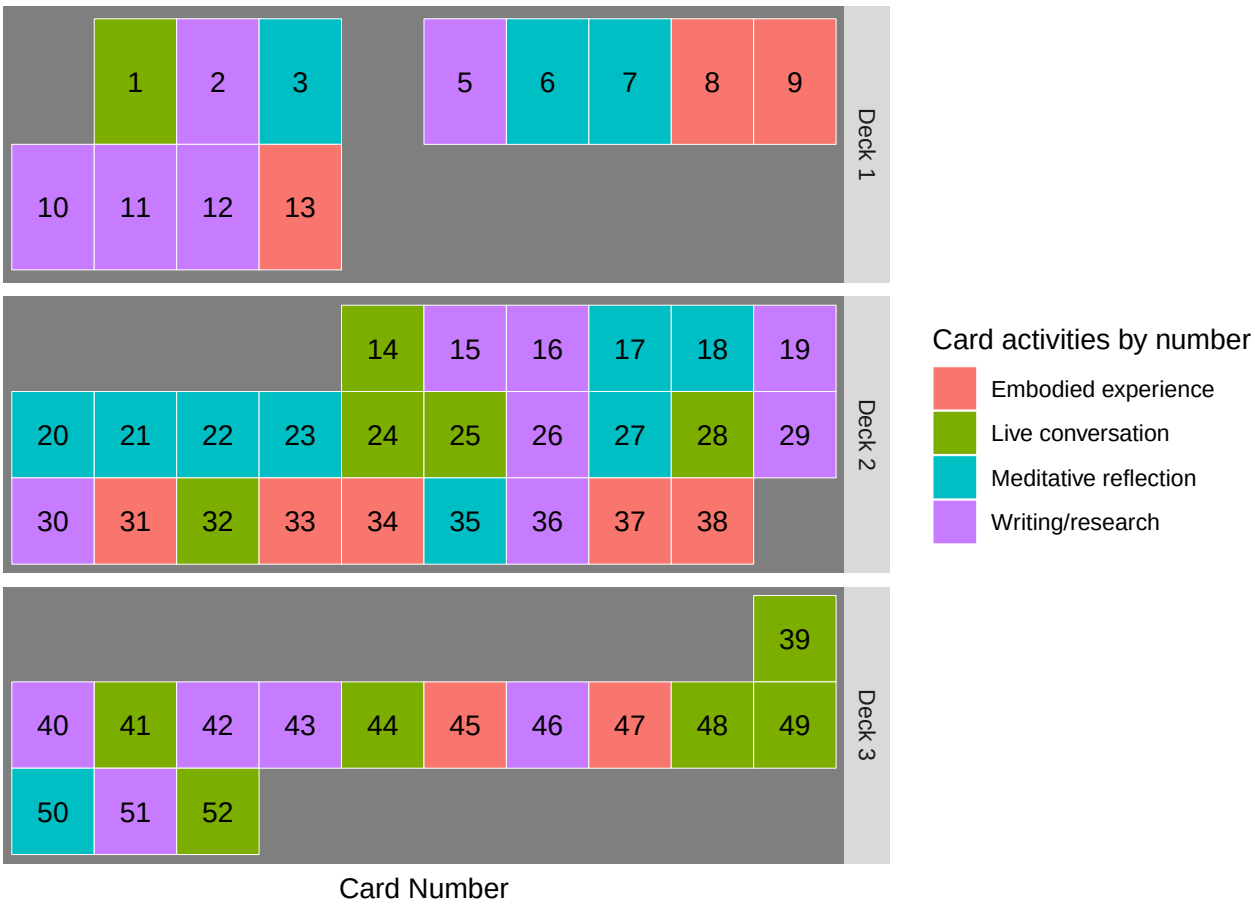

Figure 7: Card types across decks

On average, we find that players were noticeably more likely to play cards that involved live conversation. Players were about 12-13 percentage points more likely to play those particular cards (Figure 8). We find that this gap persists even if we remove the high-volume cards that begin each deck from our analysis, suggesting that this pattern is not explained exclusively by high-volume or mandatory cards being more likely to involve conversation.

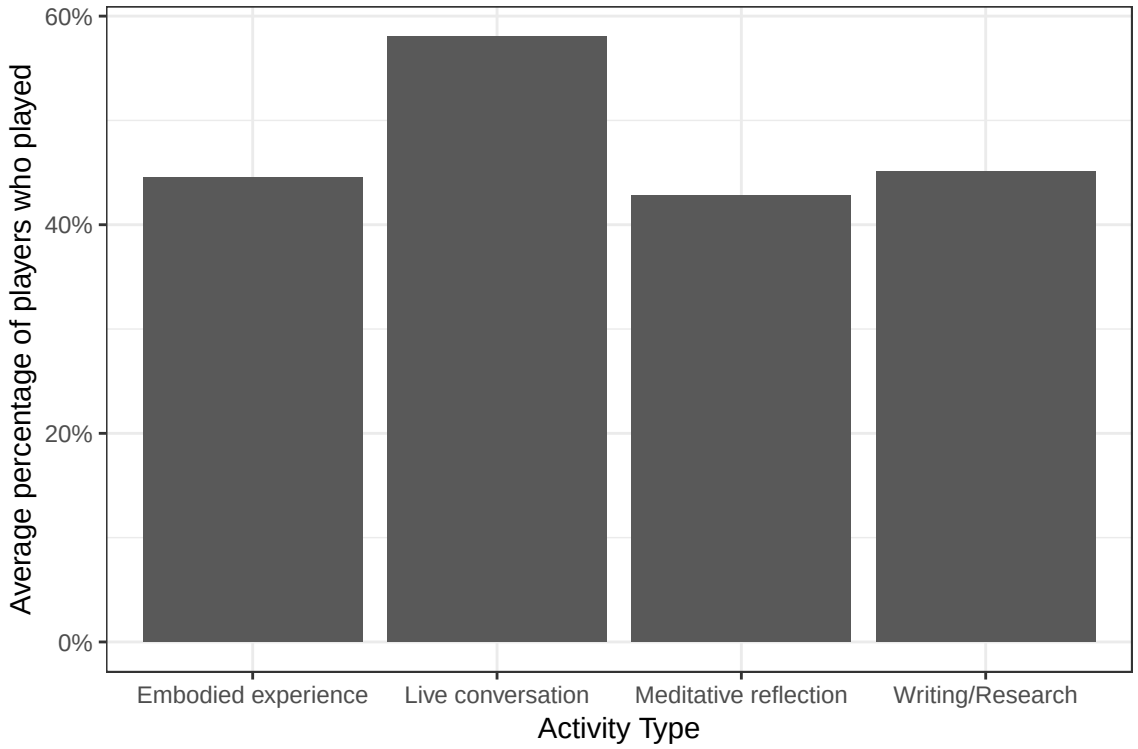

Figure 8: Percentage of players who played each card type

4 Participant drop-out and disengagement

Due to the significant time commitment involved in participating in *The End*, players were not always fully active throughout the course of the game. While the intensity of active player engagement could be expected to vary from day-to-day, we were particularly concerned about those players who would drop out or fail to check-in entirely - players who would become inactive. Built in to the game was the ability of players to choose to "rest" when necessary, notifying the actors that they would be unable to play for that day. Nevertheless, despite the access to this time management mechanism, a noticeable fraction of players would fail to contact "The End" entirely. While some of these players would re-contact on subsequent days, others eventually dropped out of the game entirely. This section analyzes drop-out and "no contact" behavior, examining its patterns over time along with any possible prognostic factors for non-contactibility.

4.1 Probability of drop-out over time

First, we look at the probability that a player does not play on a given day over the course of the four game weeks. We fit a locally smoothed regression (loess) of the indicator of no contact/resting on day and plot the estimates and 95% confidence intervals below in Figure 9.[5]

[5]This is essentially fitting a linear regression to construct a flexible estimate of the probability of no-contact for each day.

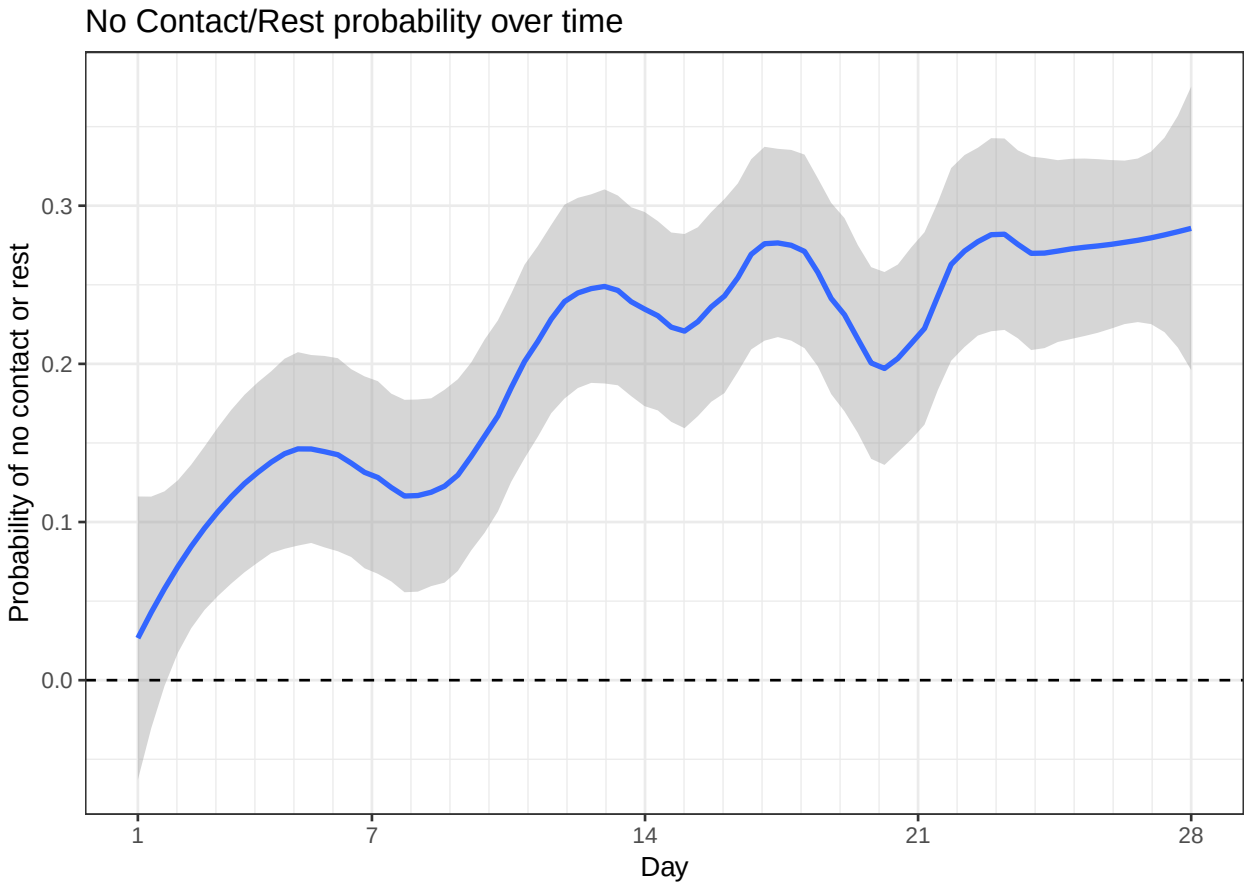

Figure 9: No Contact/Rest probability over time (57 players)

This figure gives us an estimate of the probability that any given player will choose to rest or be non-contactable in any given day. As expected the pattern is generally trending upwards as the game progresses. We then split the analysis between failure to contact and resting specifically. Figure 10 plots the estimated probability of resting in any given day while figure 11 plots the estimated probability of contact failure (excluding resting). Since the designers anticipated that players would need a break, resting is built in to the game design and therefore still reflects “correct" play.

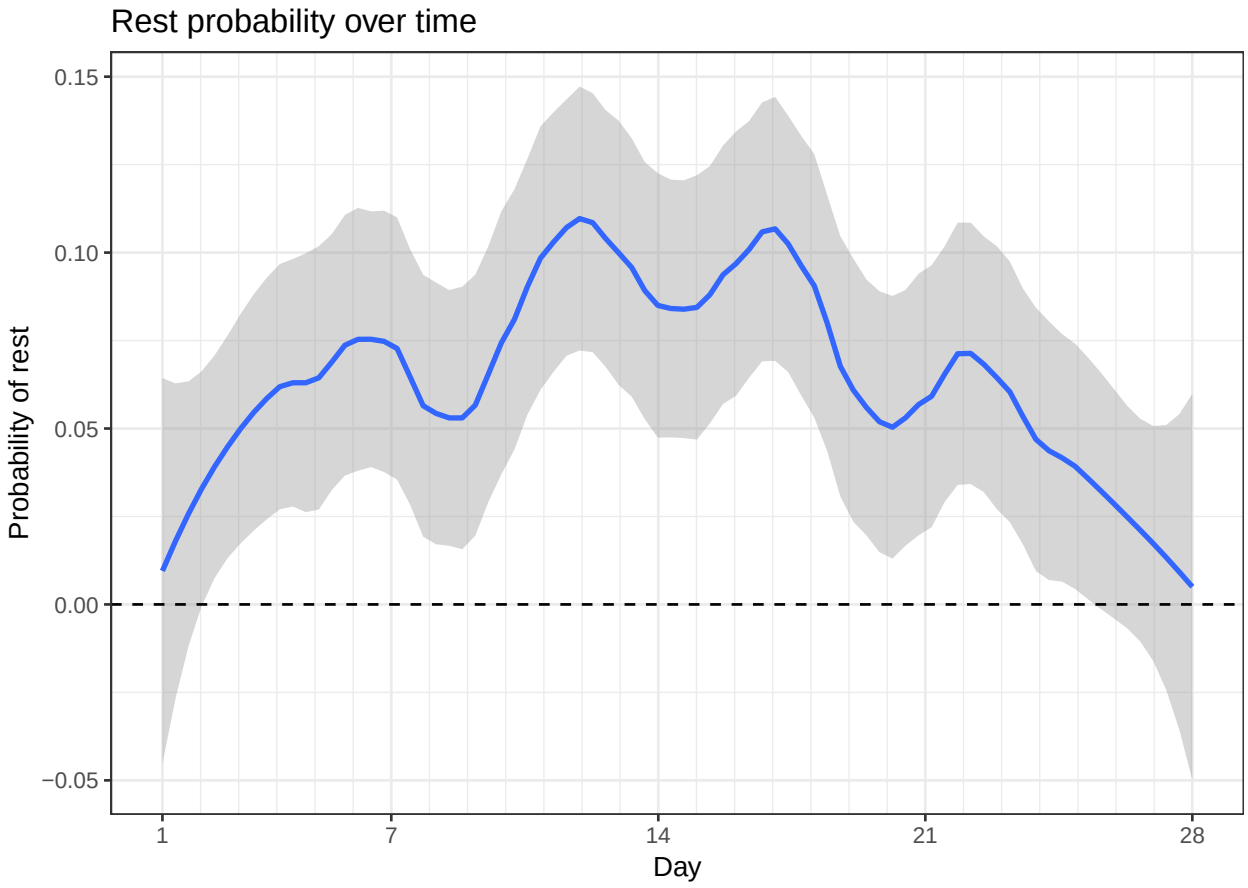

Figure 10: Rest probability over time (57 players)

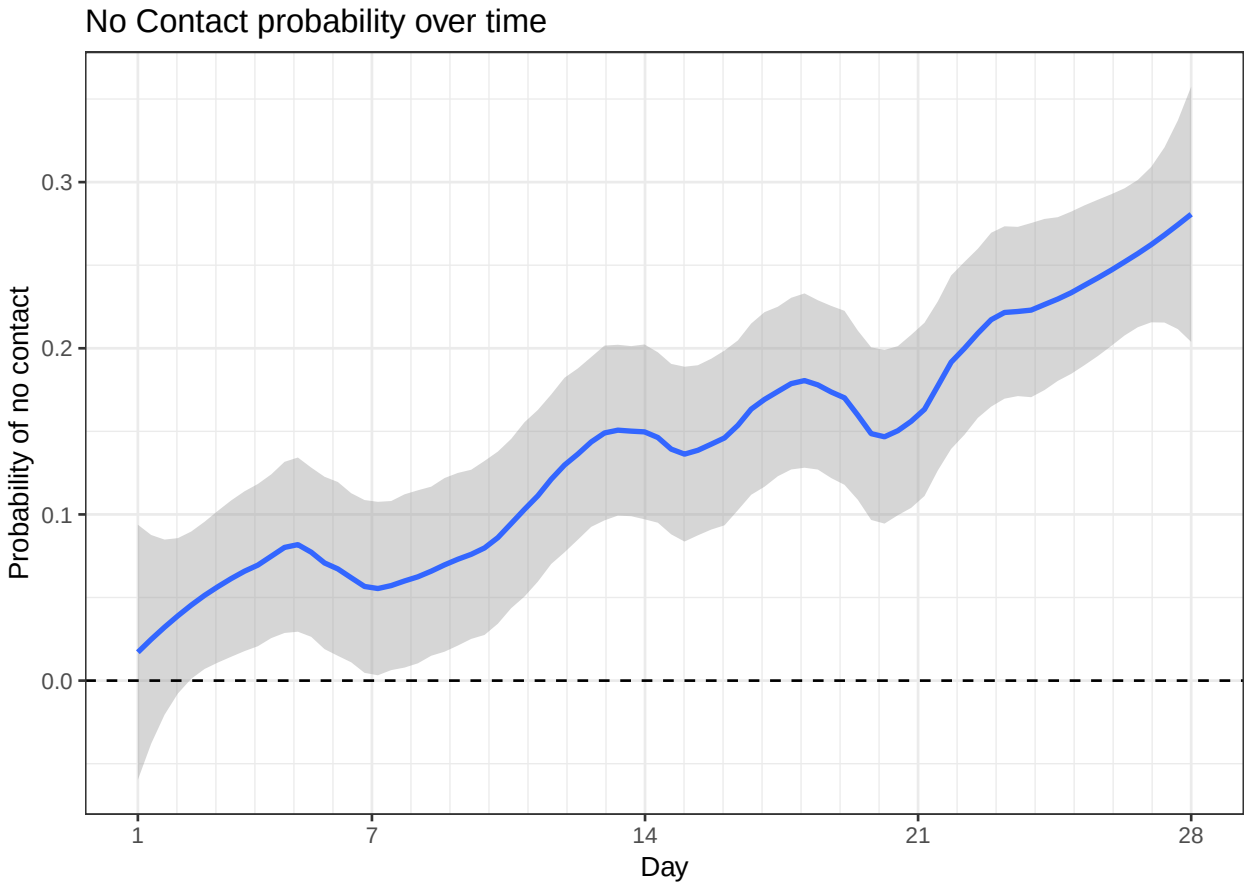

Figure 11: No contact probability over time (57 players)

We see that resting peaks in late week 2/week 3 and fades out over the last week possibly due to more ambivalent respondents dropping out entirely towards the later period. Interestingly most of the day-to-day variation is in rest behavior, while the growth in overall non-contactability is a consistent upward trend.

We evaluate whether resting and non-contactibility are connected by looking at the probability that a respondent will be non-contactable conditional on their history of resting. First, we look at the distribution of how many times respondents rest during the game. Figure 12 plots a histogram of the number of times respondents rest during the game

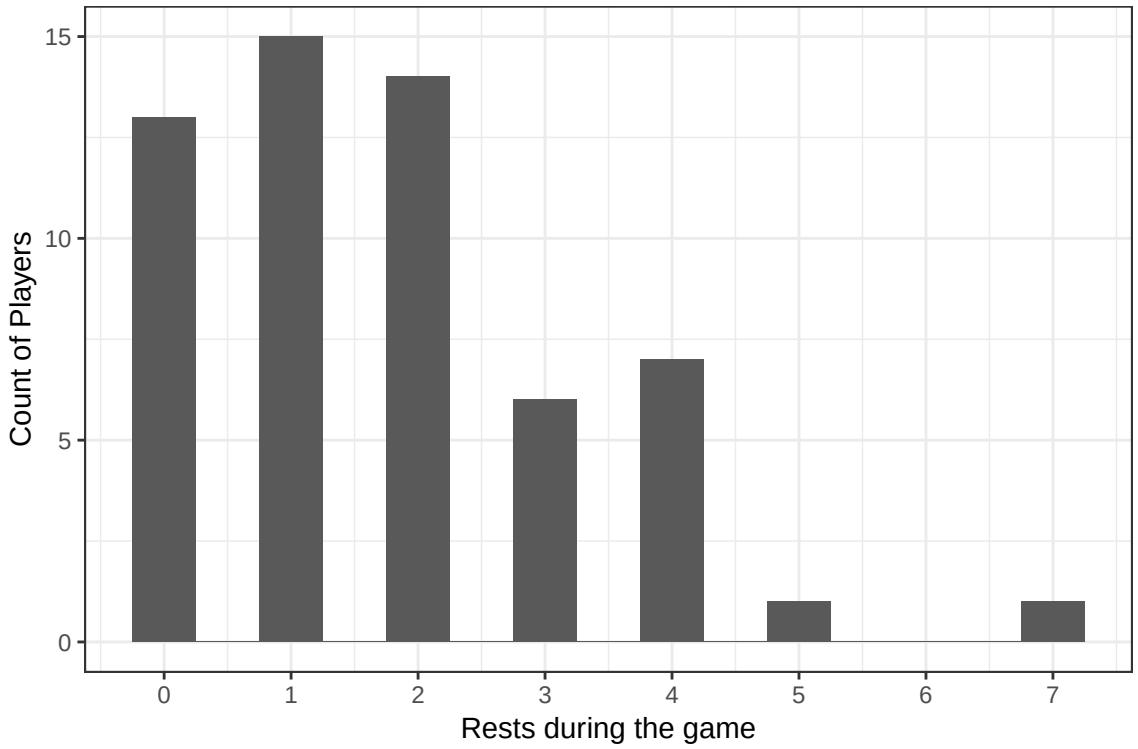

Figure 12: Distribution of the number of times players rest

Most players rested only about 0-2 times, with the modal number being 1 rest. A smaller fraction rested 3+ times, with one player going as a far as to rest 7 times during the game. Given a player's resting history, what's the probability that they will fail to be contacted in the subsequent day? We coarsen the rest count variable to six levels: 0, 1, 2, 3, 4+ and fit a regression of no-contact on the aggregated rest count and a linear time trend (to control for the fact that the cumulative number of rests is obviously trending upwards over time).

Overall, we find that players who have rested once or twice are no more likely to be non-contactable in future periods. We found a slight positive but not statistically significant relationship between past rest history and non-contactibility once we get out to 4 or more rests. But because there are very few respondents with profiles that reach that many rest periods, the data is very noisy and it is hard to draw any meaningful conclusions about the relationship. In general, those players using a few rests do not appear to be less engaged than those who are playing continuously.

4.2 Predicting non-contactibility

Are certain types of players more or less likely to be non-contactable or drop out of the game? If attrition rates vary across characteristics that are measurable at the beginning of the game, we could potentially design interventions to improve accessibility for these players. We start by looking at the demographic characteristics collected in the intake surveys. We aggregated the total number of non-contactable days for each respondent and fit a linear regression model predicting the number of non-contactable days using age and indicators for gender and race.

In general, none of the variables was found to be a statistically significant predictor of the number of days that a player is non-contactable. However, there may be some interesting underlying non-linearities. Plotting the relationship between the number of non-contactable days and age (Figure 13 suggests a visibile improvement in contactability occurring primarily around the late 30s). Aside from this possible non-linear relationship

with age, none of the measured variables in the intake stage appear to be significantly associated with whether players would subsequently be contactable.

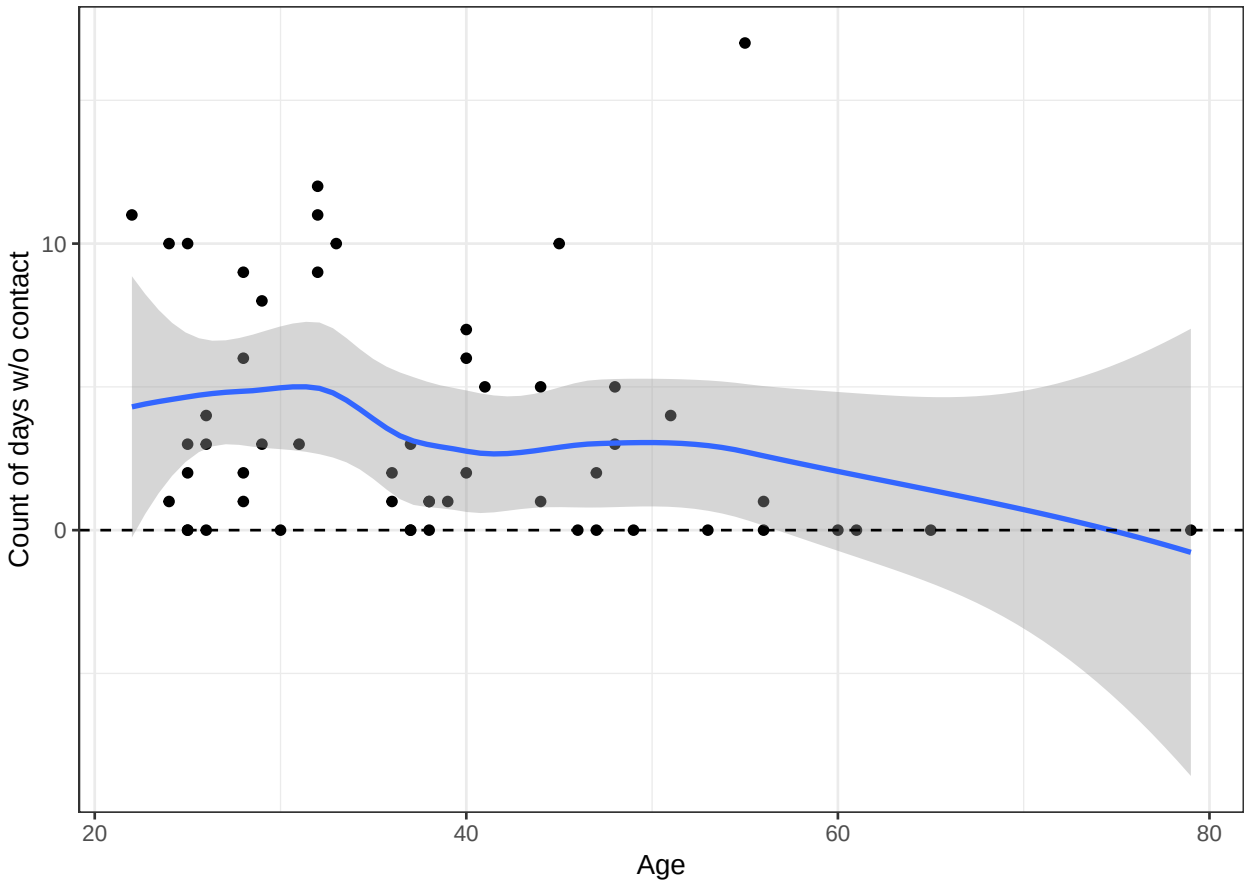

Figure 13: Non-contactability by age

During the intake surveys, players were asked a number of questions related to their history of mental health and experiences with loss. It would be useful to know whether or not these variables are predictive of any element of players' ability to play the game considering the subject matter of *The End*. We consider two separate outcomes: the total number of days that a player is not able to be reached and the total number of days that a player chooses to rest. We regress each outcome on five separate survey questions:

- (`Experienced Loss`): Have you experienced many losses in your life? (scale of 0 to 4, with 4 being "I've had many exerpiences of loss")
- (`Depression`): Do you have a history of depression? (0=no, 1=yes)
- (`Panic Attacks`): In the past four months, have you experienced panic attacks? (0=no, 1=yes)
- (`Professional Available`): Are there people or professionals you turn to when you need emotional support? (scale of 0 to 4, with 4 being "I have an excellent support system")
- (`Receive Help`): To what extent do you receive help and support from others when you need it? (scale of 1-10)

Results are plotted in figures 14 and 15. In each plot, the coefficients denote the estimate of the expected change in the outcome (number of days) for a one unit increase in the independent variable, holding all other variables constant. We omit from the plots the intercept for each regression.

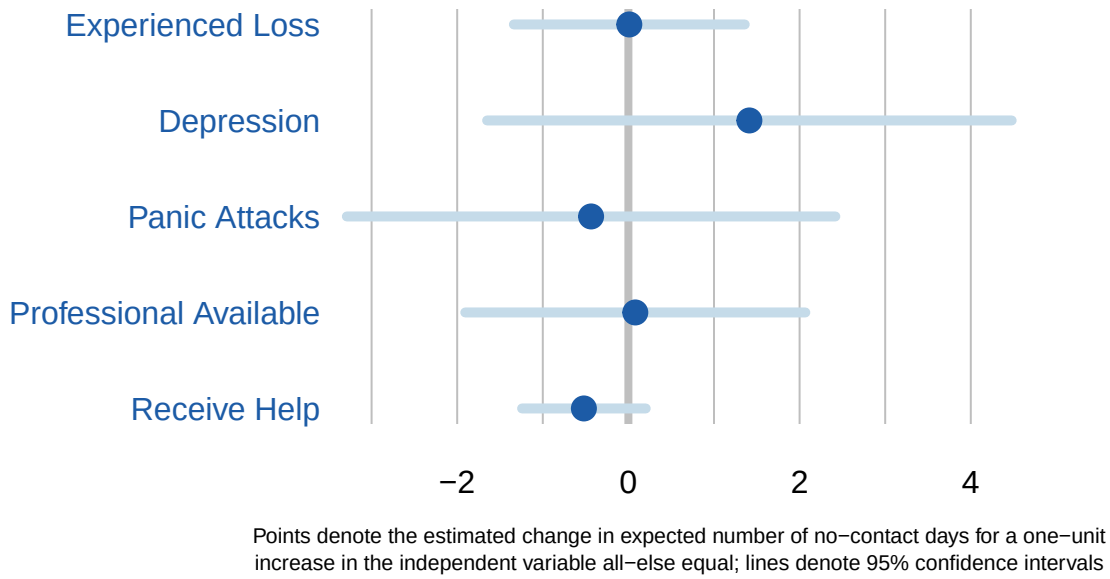

Figure 14: Regression of number of no-contact days on survey questions (red = p < .05)

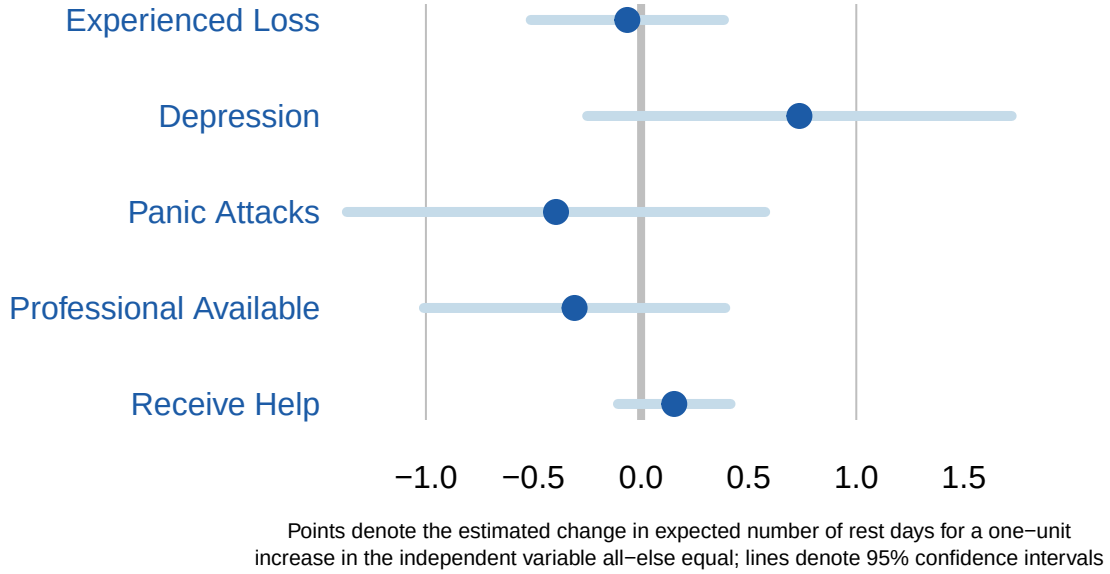

Figure 15: Regression of number of rest days on survey questions (red = p < .05)

The sample size is relatively small (only 55 total players) and therefore it is difficult to definitely conclude that there are definitely no differences in outcome across different levels of the survey responses. Nevertheless, the results do not point to any strong relationship between ability to play the game and reported mental health status. None of the estimated regression coefficients on the six survey variables reaches conventional levels of statistical significance across all five regressions.

5 Cards and variation in play

One very important question in understanding both the design and the impact of *The End* is whether different cards had different effects on the players who experienced them. While unfortunately both the non-random assignment of cards to players and the small sample size make any inferences here quite difficult, it is useful from a descriptive standpoint to see whether certain cards elicited certain types of reactions from players, particularly if they are unexpected. This section analyzes player activity with respect to card choice along with the impact of certain card types on player behavior

5.1 Card type and non-contactability

A potential question related to players' frequency of play is whether players who play certain cards are associated with greater or lesser attrition. This is a tough question to answer statistically since we have many cards in the deck. Just by chance, if we are testing many cards, one card could be by chance strongly associated with the outcome of interest. One strategy for limiting this "multiple comparisons" problem is to use a variable selection method. A commonly used variable selection tool is known as the "lasso" which fits a linear regression model with a penalty term. This penalizes solutions that have large absolute values of the coefficients, and intuitively the form of the penalty "shrinks" the smaller coefficients towards zero.

We first fit the lasso regression, and look at which cards have estimated non-zero coefficients. We then take the non-zeroed out cards and put them in an ordinary least squares to obtain unbiased coefficient estimates along with standard errors (an approach known as "Post-LASSO"). Figure 16 plots the results of this second post-Lasso stage.

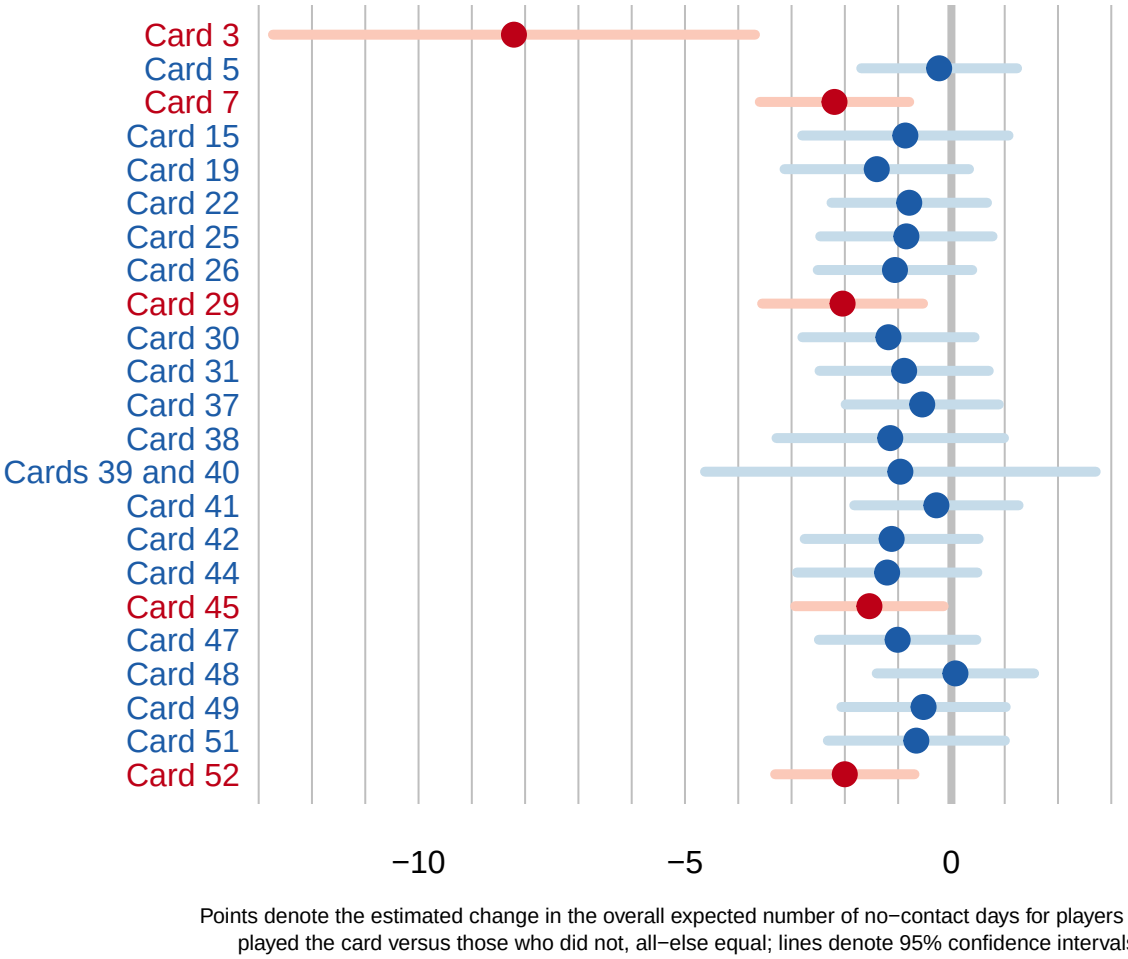

Figure 16: Regression of number of non-contactable days on whether players played each card (red = $p < .05$)

In general, and as expected, players who play any given card have fewer non-contact dates than those who do not (since, by definition, non-contacted players are not playing a card). So it's important to take with caution any interpretation of these observed associations as causal. Among these cards, however, we have a few interesting associations. Obviously, players who played Card 3 had, on average, many fewer non-contacted dates. This likely reflects the fact that players who are less active early in the game are also less likely to be active throughout the course of the game - early behavior predicts future behavior. Among the other early choices, players who played card 7 were on average more active throughout the game. Cards 29 and cards 45 also predict greater activity and less non-contactability.

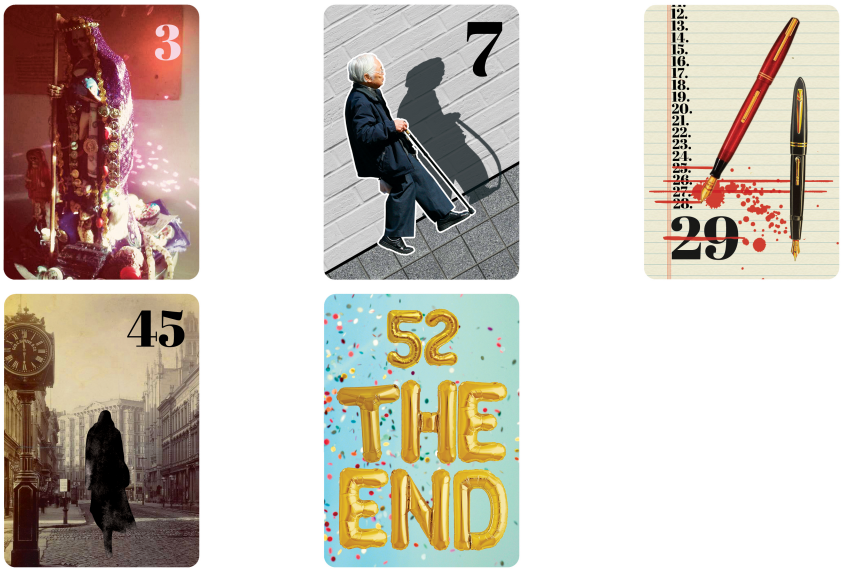

Figure 17: Cards 3, 7, 29, 45, 52

5.2 Player attributes and card choice

Within the game, there are many moments where players have the ability to choose between playing certain cards. In the early stages of the game, players are given constrained choices. This allows us to evaluate whether certain players are more likely to select certain types of cards.

Figure 18: Cards 5 and 6

The first choice players are presented with is the option of selecting card 5 or 6 after playing card 4. The majority of players chose 6 over 5. It turns out that not that many of the demographic covariates predict well whether players choose card five or six, but one of the intake survey questions does. Players who indicated

that they have "have an excellent support system" and have professional help available are about twice as likely to choose card five over six when offered the choice (Figure 19).

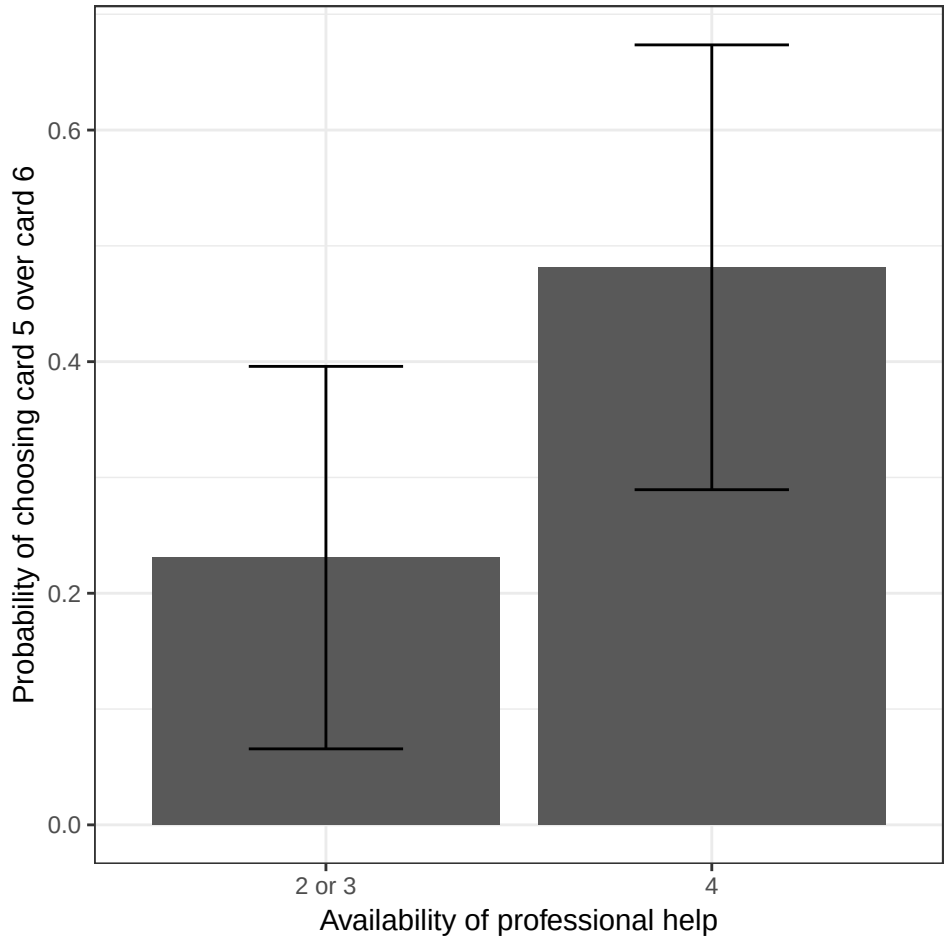

Figure 19: Probability of choosing Card 5 by availability of 'Professional help'

Outside of five and six, players' choices are a bit more open-ended and not necessarily mutually exclusive. Moreover, not all players necessarily get the same choices since they are in part guided by the actors. We run into a strong possibility of finding spurious findings by comparing all possible cards with all possible covariates. The cards in the first deck, however, provide a somewhat more constrained test as most of the deck is heavily guided and players have limited choice at the begining. The other card we find has a strong association with demographics is card number 9. We find that men are on average much less likely to play card 9 (by a factor of about 3) (Figure 21)

Figure 20: Card 9

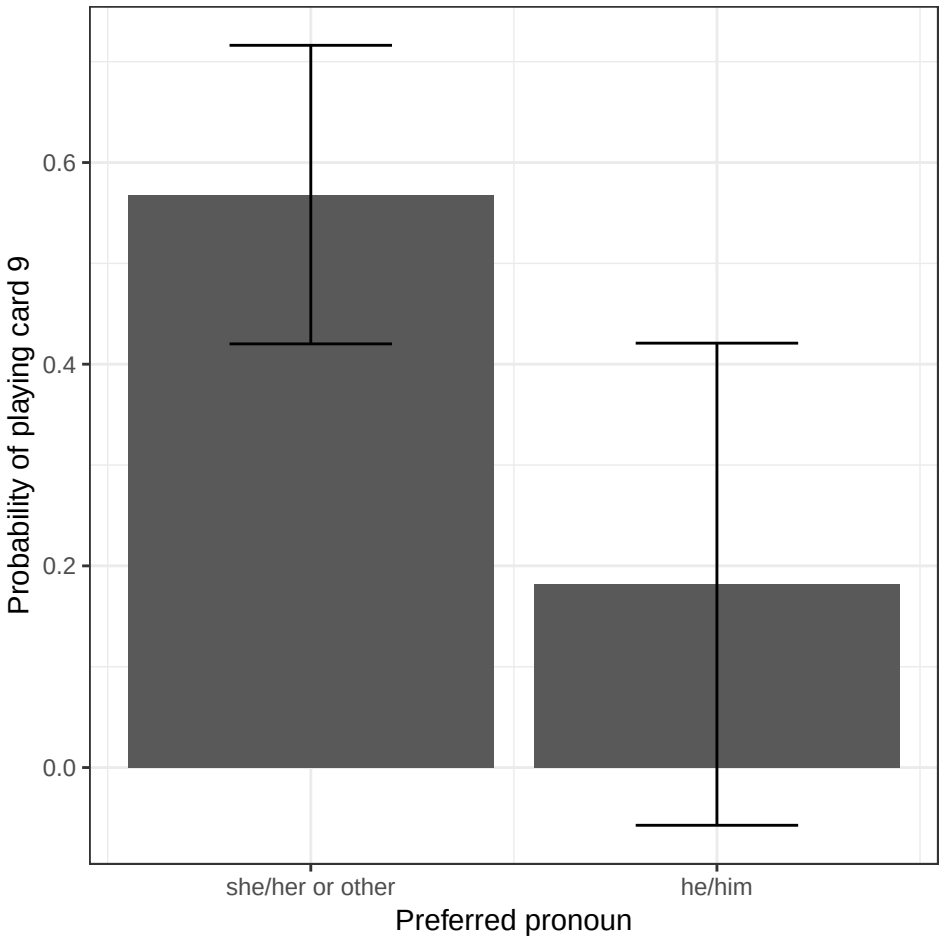

Figure 21: Probability of choosing Card 9 by gender (lines denote 95% asymptotic CI)

5.3 Card choice and follow-up survey measures

Does experiencing a particular card cause players to improve in their assessments of their own lives? We evaluate whethere there is an association between playing of certain cards and changes in some of the flourishing metrics that were measured. We focus specifically on the questions related to belief in a purposeful and meaningful life and general happiness. A total of 39 respondents answered these questions pre- and post-experience. We examine whether the change in individual metrics before and after the game is associated with playing or not playing any one of the 52 cards. Again, we use the lasso to fit a regression of the within-respondent change in happiness and purposefulness on the full set of card play indicators. Notably, without the regularization, this regression would be impossible as we have more variables than observations!

With respect to player happiness, it appears that only one of the card play variables has a non-zero coefficient when fitting the lasso regression: Card 32

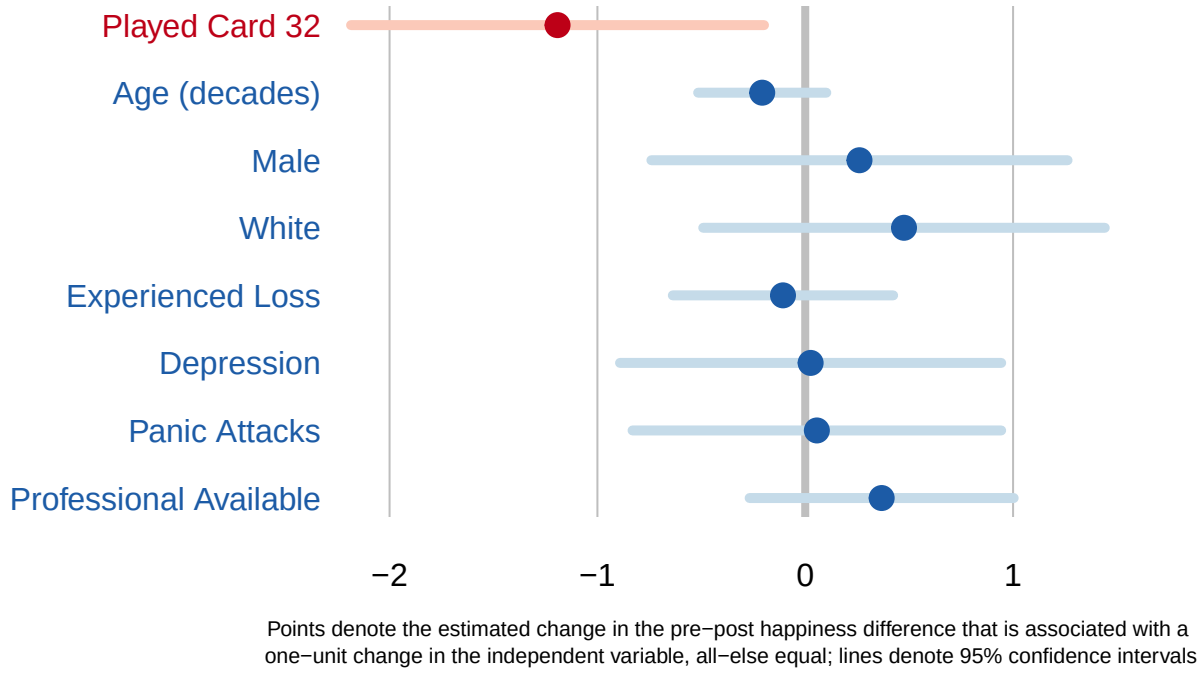

Figure 22: Regression of change in self-reported happiness on whether players played card 32 + controls (red = p < .05)

On average, players who played card 32 had a lower pre-/post- change in their reported happiness - in other words, they were less happy than those who did not play card 32, controlling for baseline happiness levels. On average, players who didn't play 32 tended to be happier after the game relative to their baseline. Those who played 32 stay at about the same level afterwards. The overall magnitude of the effect is quite sizeable as well: a nearly one standard deviation change in the happiness measure.

This appears to hold even when we control for other factors like age, gender, race, along with some of the other mental health indicators (Figure 22). Again, this relationship is not necessarily causal as it may be that players who were more likely to have a decline in happiness between start and end of the game were also more likely to play card 32. Nevertheless, it appears that there is something associated with this card (either the card itself, or a selection process for the types of players who played 32) that influenced players' self-reported happiness. This observed relationship is puzzling but interesting given that Card 32, titled "One Last Wish," is not obviously designed to evoke a negative response. The card itself tasks participants with sending a "blessing or a wish to the world to be bestowed after you are gone" and in part asks them to contemplate what they hope for the future. It is worth noting that this is one of the few cards that takes a

more “macro" level view of life beyond the context of the player. In general, most cards focus on personal relationships when thinking about life after death while this card is more general in its framing.

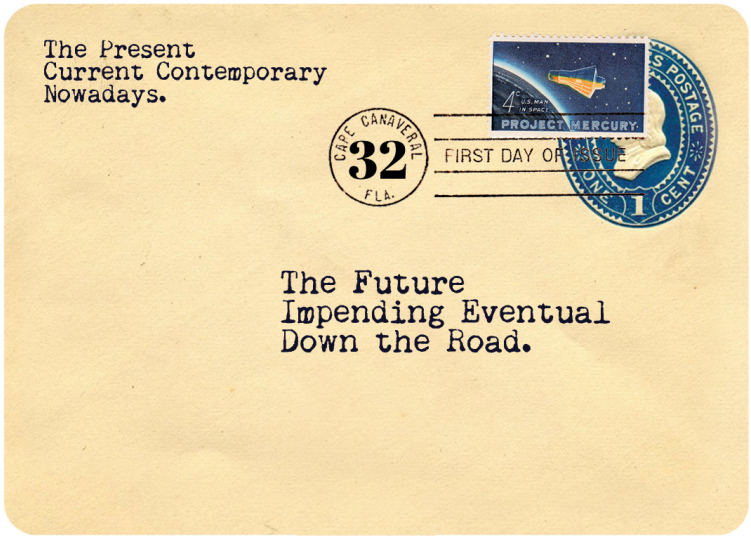

Figure 23: Card 32

On the purposefulness measure, the lasso sets all coefficient values to 0. Given the small sample size, this is not unexpected. There isn't that much data to go on here, unfortunately, but it is not clear that there is a single card that is strongly associated with changes in self-reported purposefulness ratings.

5.4 Activity levels by card

While above we considered the propensity of players to select and play each card, in this section we examine the overall amount of activity associated with each of these cards. First, we condense the data on the number of players who played each card into a simple heat map (Figure 24).

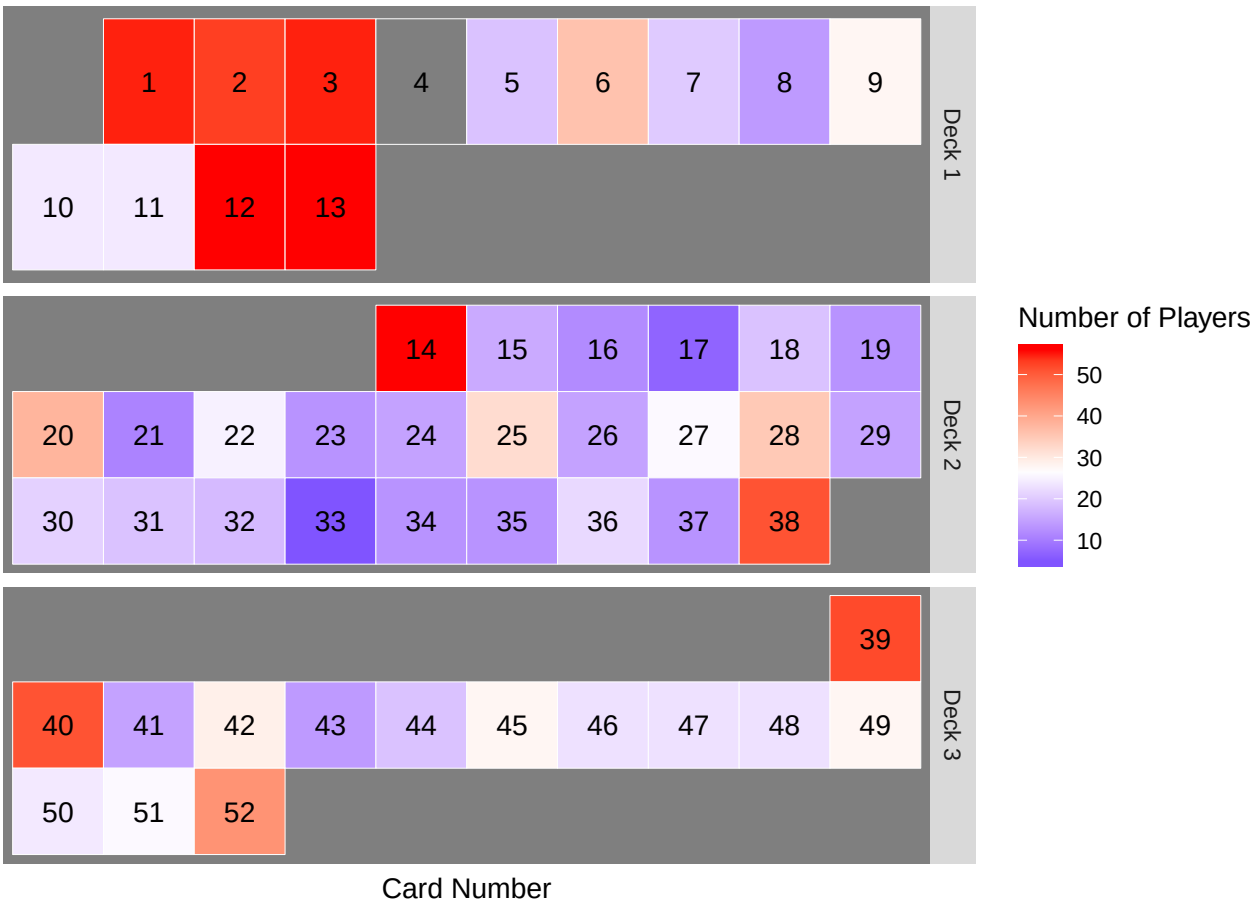

Figure 24: Number of players playing each card

Cards 1, 2, 3, 12 and 13 in Deck 1, 14 and 38 in Deck 2, and 39 and 40 in Deck 3 are required by gameplay. We then plot the average number of messages sent by players (among those who played) each card (Figure 25).

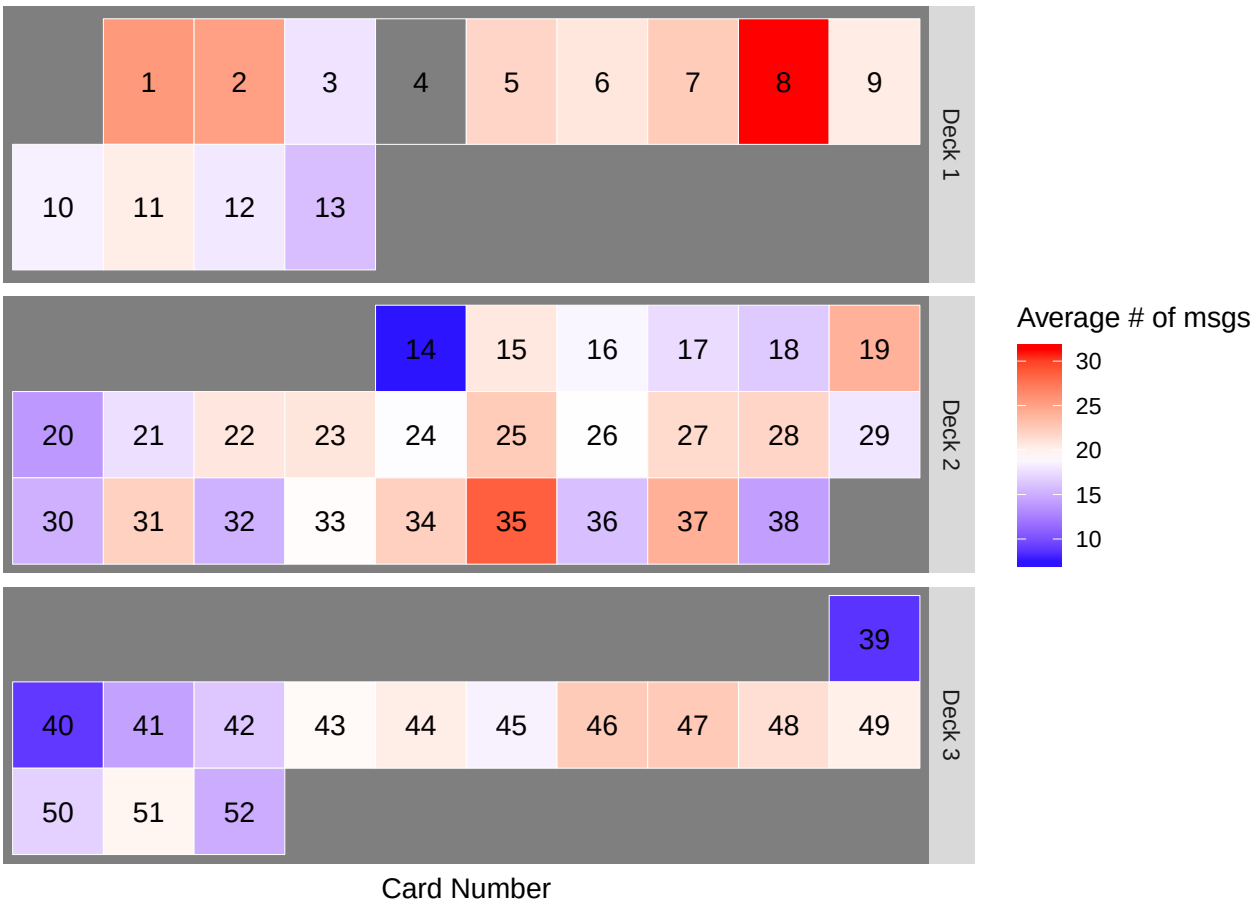

Figure 25: Average number of messages sent by players

We also consider the average message length in number of words for each card (Figure 26)

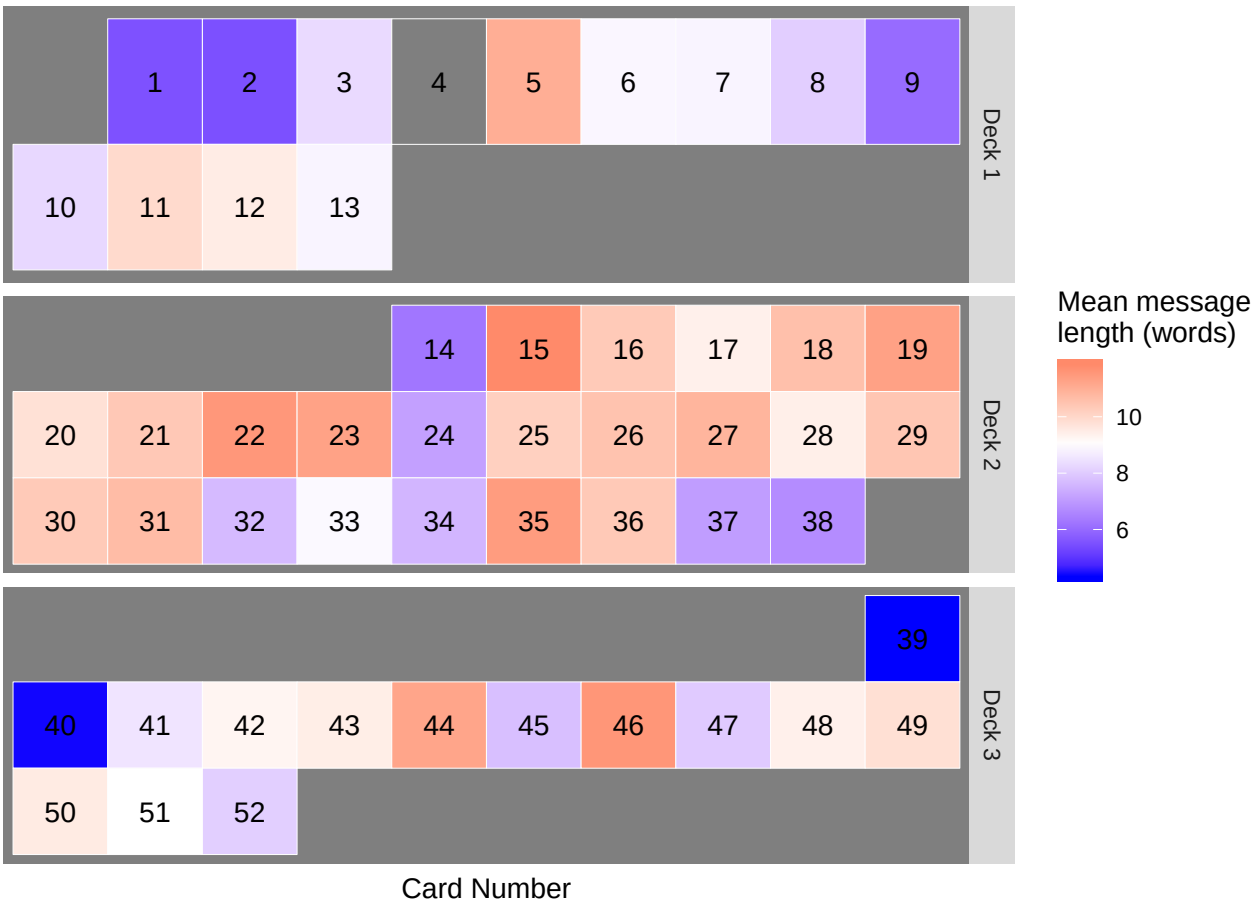

Figure 26: Average message length per card

Finally, we plot a combination of number of messages and message length - the average number of words written by players who played each card (Figure 27). Cards 14, 39 and 40 have conversations that take place in Google docs instead of through text message, which likely explains the low word density for these cards.

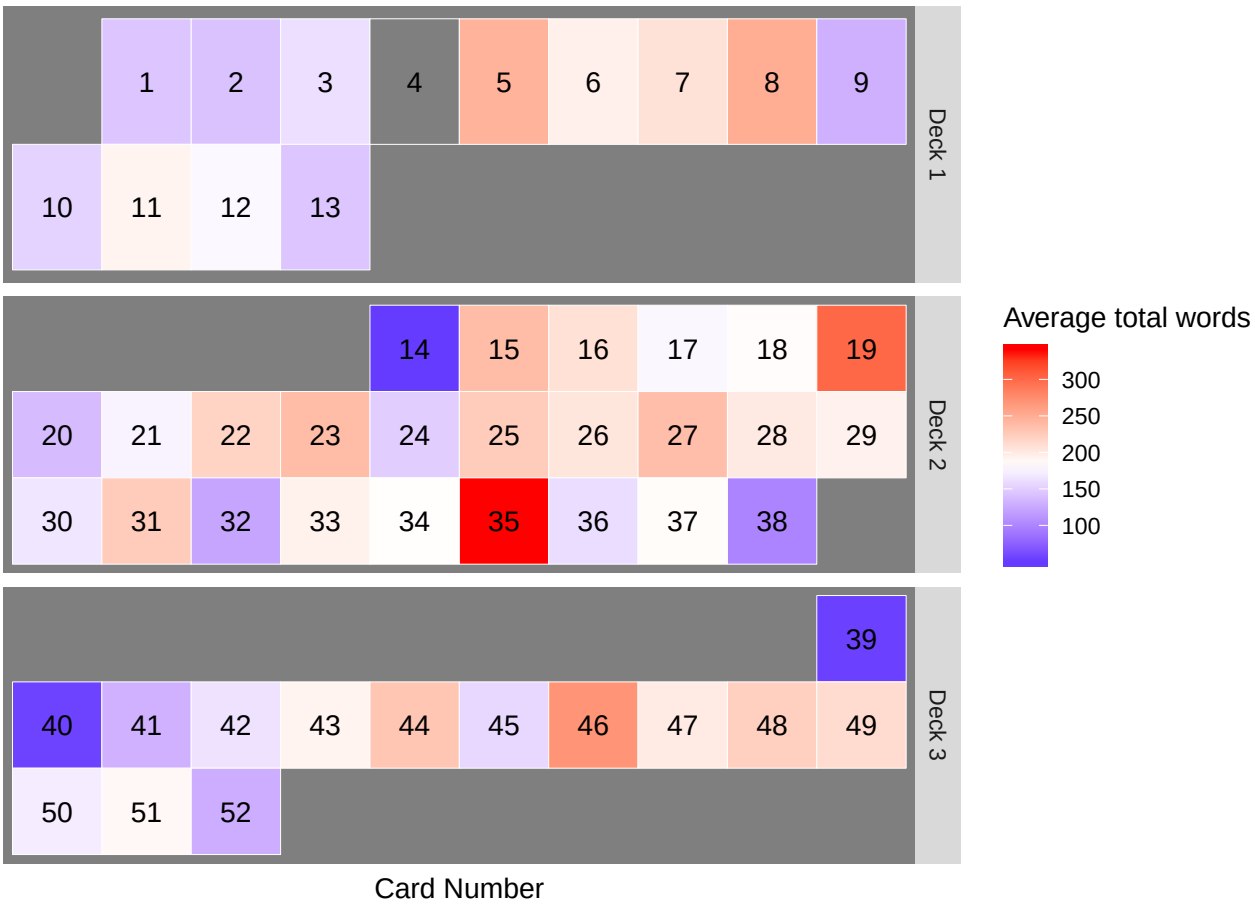

Figure 27: Average number of words written by players per card

One card that stands out in particular for its high levels of engagement (both in terms of number of responses and message length) is Card 35. This card, titled "Marking Time" tasked players with imagining a historic marker left behind in their physical place. It asked them to draw a rough sketch of this placard and text in that sketch or a word that captured the experience. Interestingly, this particular card did not further prompt the player to write anything extensive yet it appears to have generated significant back-and-forth engagement between the players and actors.

Figure 28: Card 35

We also aggregated these metrics for each of the four activity types and plot the average of the per-card message lengths and per-card number of messages sent (Figure 29). Interestingly, cards involving meditative reflection see somewhat longer messages on average compared to the "embodied experience" and "live conversation" cards. There appear to be slight, but statistically negligible differences in the number of messages sent across the four categories. As we expected, those cards involving "live conversation" see fewer messages and less depth of interaction in general as they involve some activity conducted outside of the text message channel.

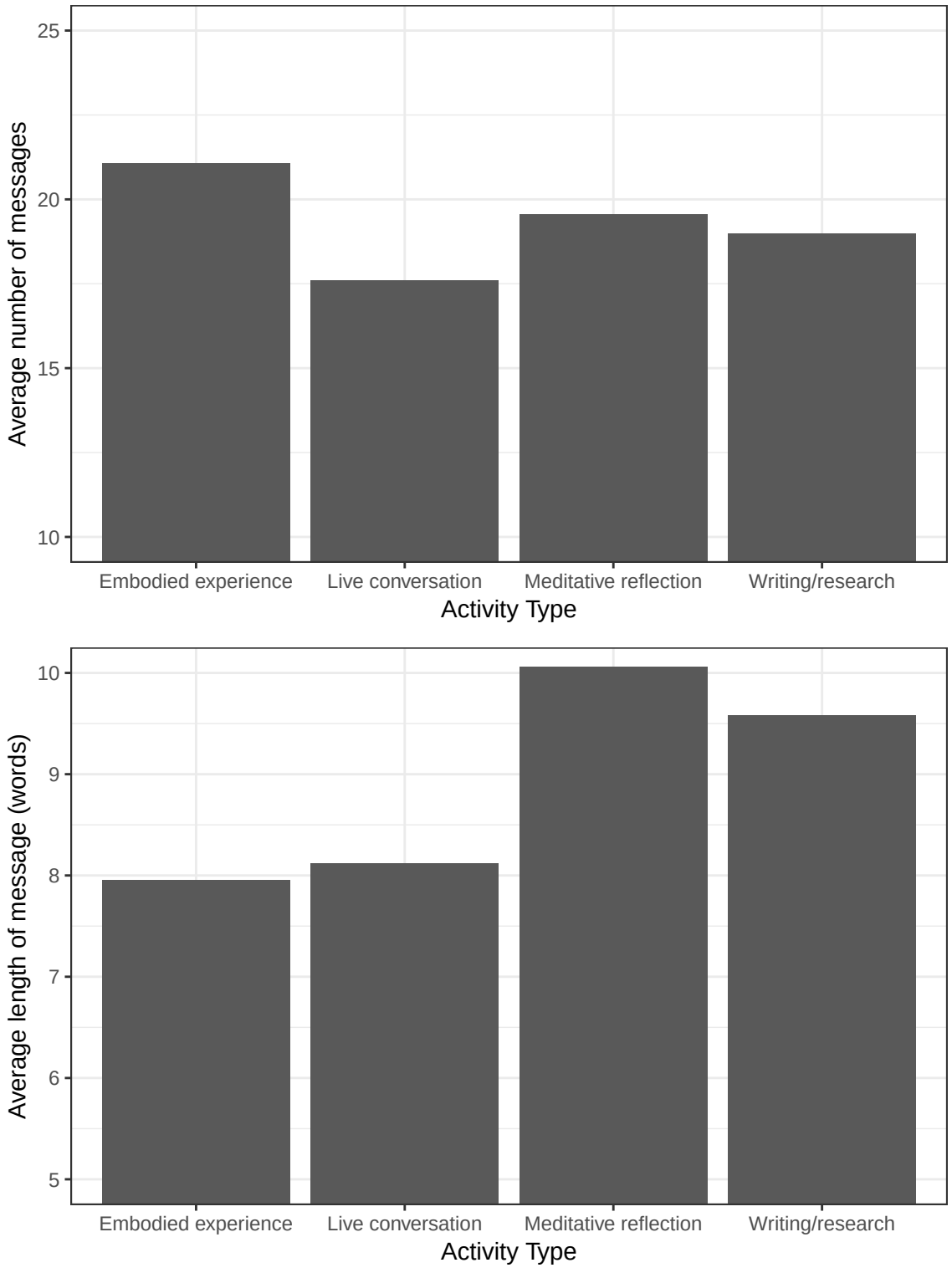

Figure 29: Number of messages and average message length by activity type

6 Measuring engagement with human-coded responses (player maps)

As the final component of the analysis, we had human coders analyze the player responses to obtain a more detailed measure of whether players' messages indicated shallow or deep degrees of engagement with the game. We refer to this as the *player map*. Codes were created by Adrienne Mackey, lead artist of the game, to assess four key values that the creative team used as guides in running *The End*. Human coders could assign each message to the four non-mutually-exclusive categories: "Stepping Into The Magic Circle" (Category 1) which denotes the necessary suspension of disbelief and acceptance of the game mechanics and rules that is a pre-requisite for engaging with the game, "Displaying Vulnerability and Bravery" (Category 2) which captures players' willingness to accept or engage with new, surprising or difficult experiences, "Reflecting On and Defining Personal Values" (Category 3) which represents retrospective or inward looking conversation regarding players' attitudes and/or beliefs, and "Taking Agency and Undergoing Transformation" (Category 4) which involves players making concrete steps for action in the present or future. Overall, the designers of the game considered the four categories to form a pyramidal structure. Category 1 was expected to comprise the largest share of interactions. Engagement with the game would build into Category 2 - engagement with new experiences, which in turn would build into Category 4 - reflection. Category 4 was considered the apex of this pyramid, the highest level of participatory behavior.

Figure 30 plots the aggregate frequency of each of these message types over time. We see that, of messages assigned a code, the vast majority fall into the first category while the action/future-oriented Category 4 has the fewest, matching expectations regarding the relative prevalence of different types of codes. It is notable to focus in on the changes in Category 4 over time. In Figure 31, we plot the total number of messages in each day that were coded as part of Category 4 along with a smooth regression estimate layered over the raw counts. We see almost tripling in the total number of such interactions from week 1 (roughly 0-10) to week 4 (20-30). Consistent with the aims of the artists to promote players' agency, we see more interactions from players that exhibit action-oriented thinking as the game progresses.

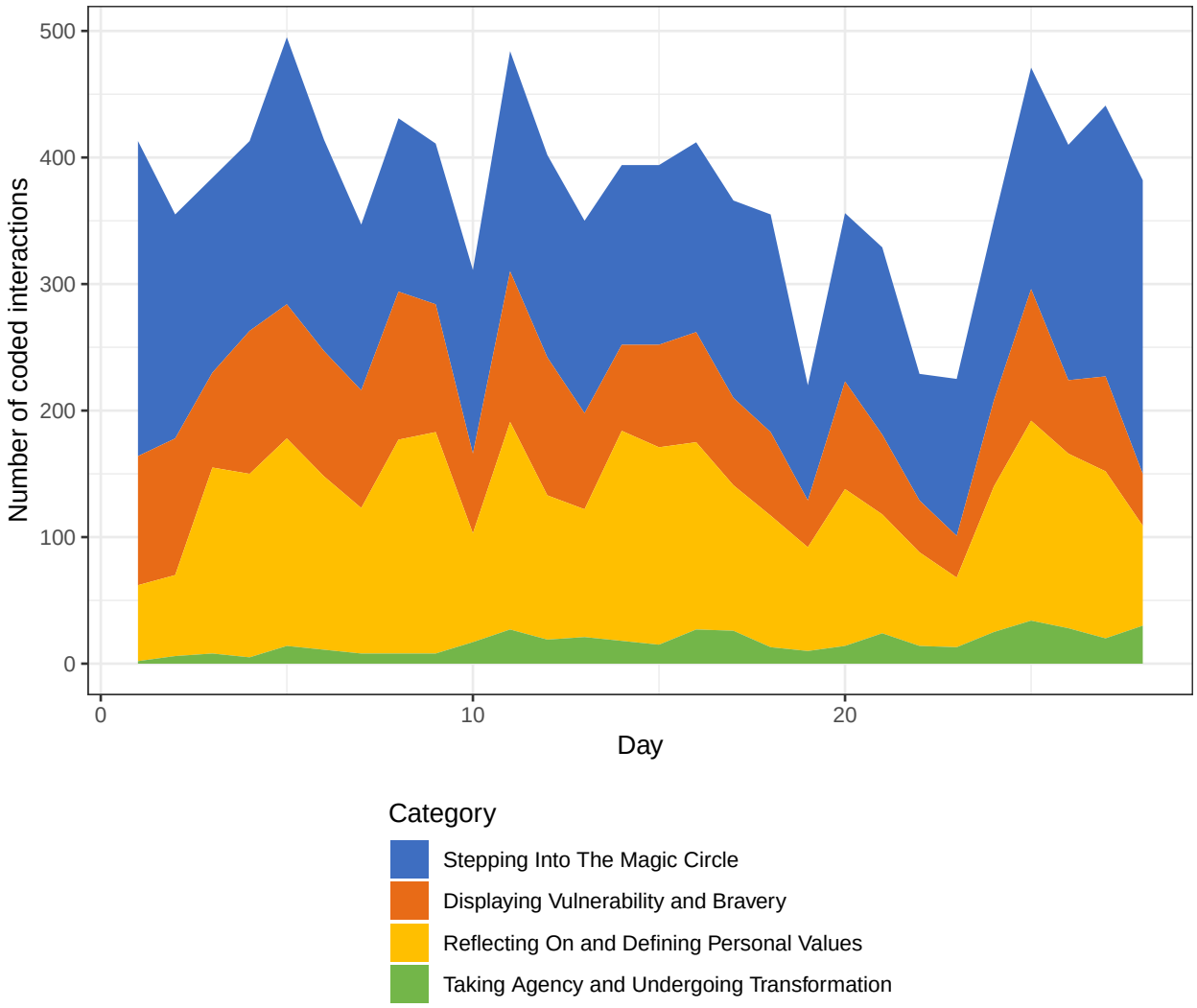

Figure 30: Day-by-day number of coded player map events

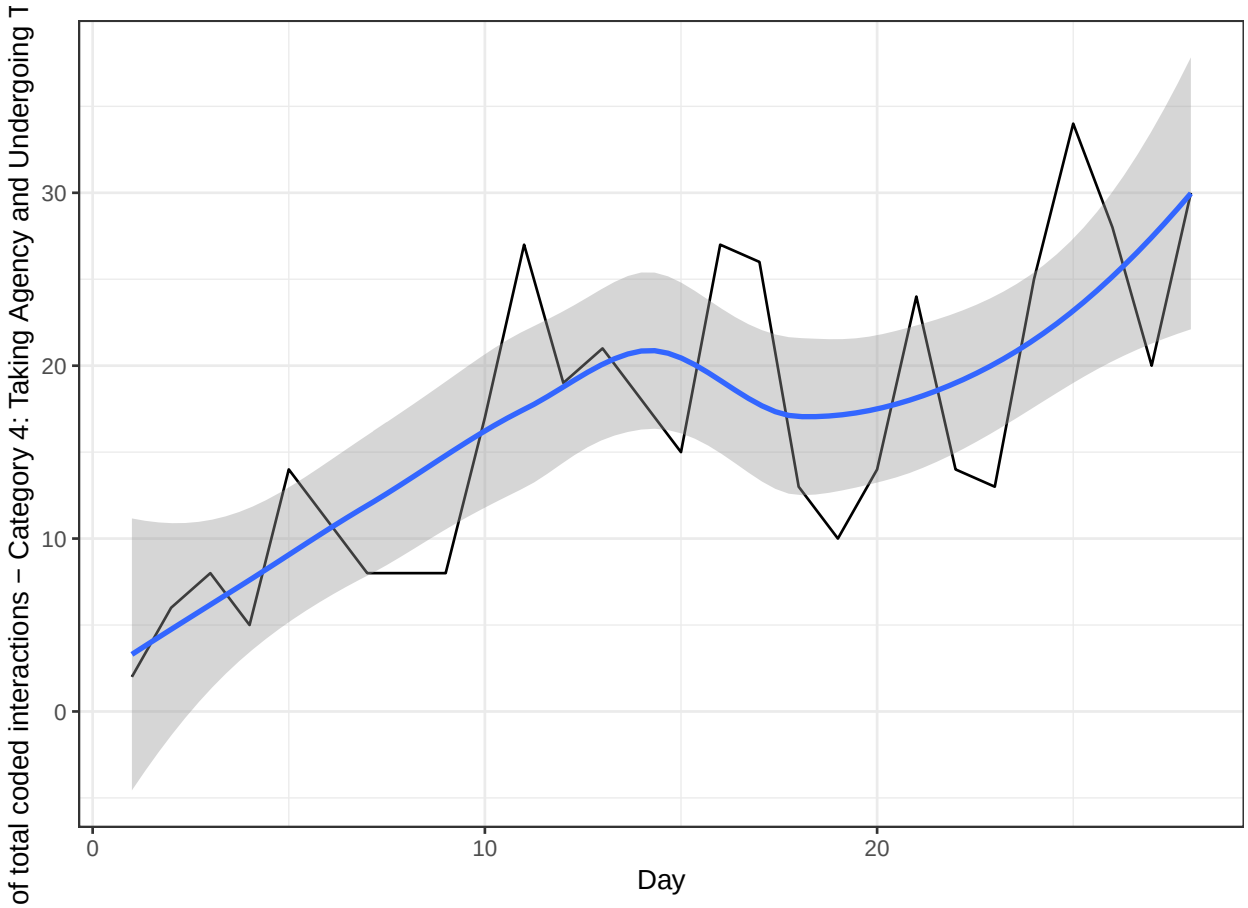

Figure 31: Day-by-day number of coded player map events - Category 4

6.1 Predicting player map behavior with survey responses

Much as we did with player responsiveness behavior, we were interested in assessing whether the mental health intake questions were predictive of whether players would have *substantive* engagement with the game as measured by the player map codes. We again regressed measures of the number of category 1 - 4 messages written by each player divided by the number of days in which they were active on the five intake questions:

- (`Experienced Loss`): Have you experienced many losses in your life? (scale of 0 to 4, with 4 being "I've had many experiences of loss")
- (`Depression`): Do you have a history of depression? (0=no, 1=yes)
- (`Panic Attacks`): In the past four months, have you experienced panic attacks? (0=no, 1=yes)
- (`Professional Available`): Are there people or professionals you turn to when you need emotional support? (scale of 0 to 4, with 4 being "I have an excellent support system")
- (`Receive Help`): To what extent do you receive help and support from others when you need it? (scale of 1-10)

Results are plotted in Figure 32 and results for the total number of *any* coded messages appear in Figure 33

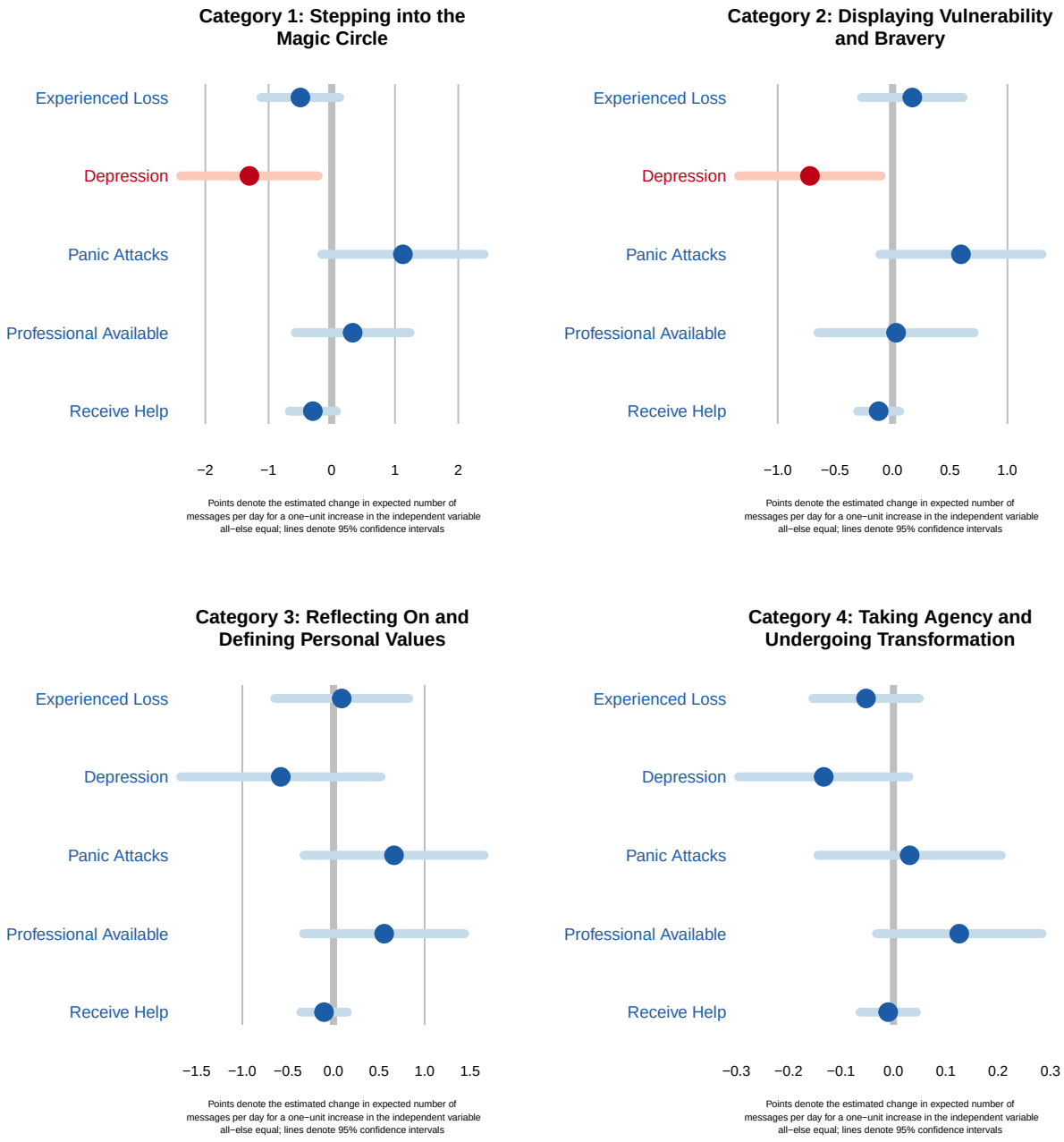

Figure 32: Regression of number of category-labeled messages per active day on survey question responses (red = $p < .05$)

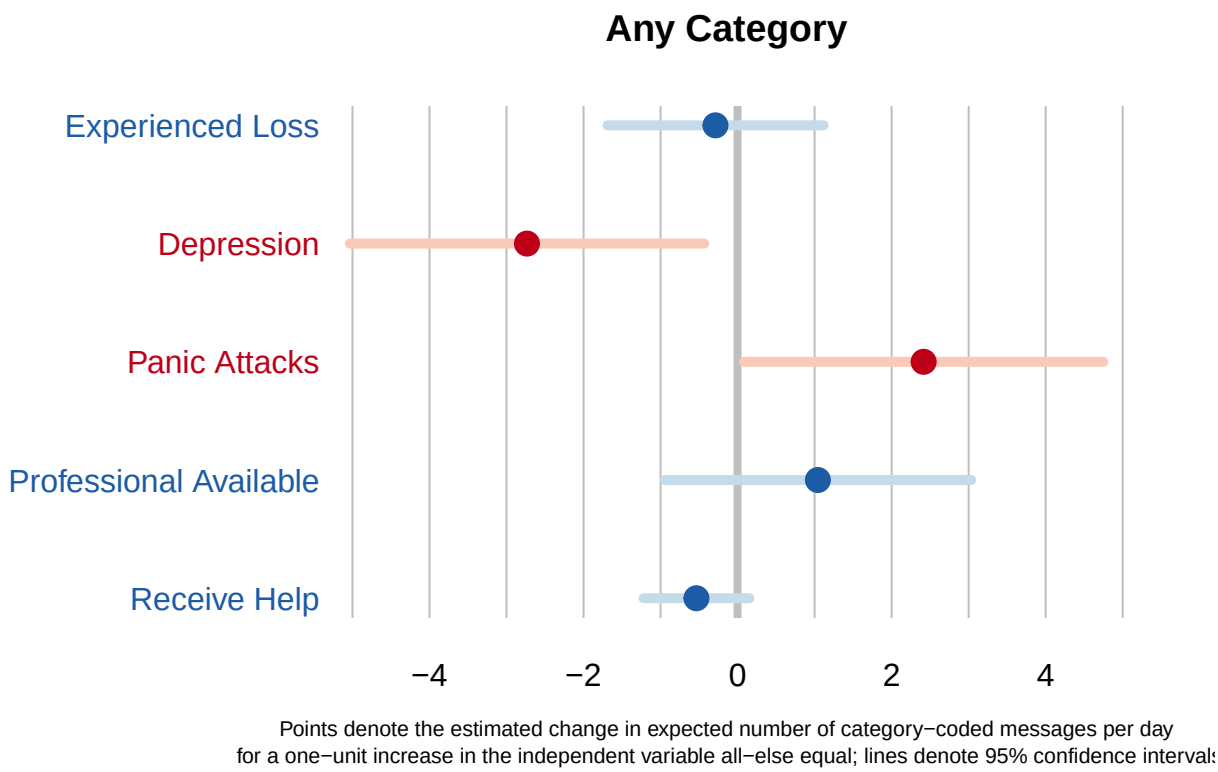

Figure 33: Regression of number of coded player map messages of any category per active day on survey question responses (red = $p < .05$)

The most striking and consistent pattern is that respondents indicating a history of depression were on average about 1 fewer message per active day in categories 1 and 2 significant at $p < .05$). In general, all-else equal, respondents with a history of depression have about 3 fewer per-day interactions that fall into one of the four categories. The results suggest that while mental health history may not be predictive of how *often* players are active or inactive, they do suggest that players with a history of depression in particular have a slightly more difficult time entering/engaging with the game at its initial levels.

6.2 Cards and player maps

Which cards are most heavily associated with higher or lower counts for each of the four engagement categories? Figures 35 and 36 visualize the patterns in the average number (per-participant) of messages assigned to one of the four player map engagement categories for each card. Among other interesting outliers, Card 8 has the most Category 1 engagement while Card 22 has the most in Category 2. Cards 23 and 26 are particularly good at eliciting Category 4 interactions. Notably, these include actionable planning as part of their prompts, suggesting that when players are tasked to create a plan of action, they do indeed do so. Consistent with our results on raw word count and message count in previous sections, Card 35 has an extremely high-level of player engagement among those who played that card.

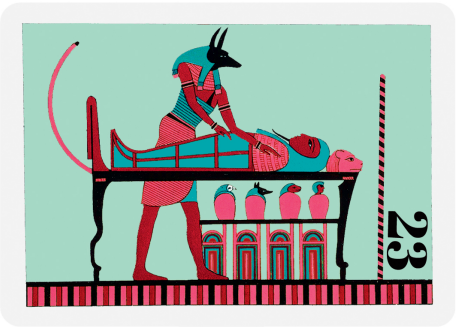

Figure 34: Cards 8, 22, 23, and 26

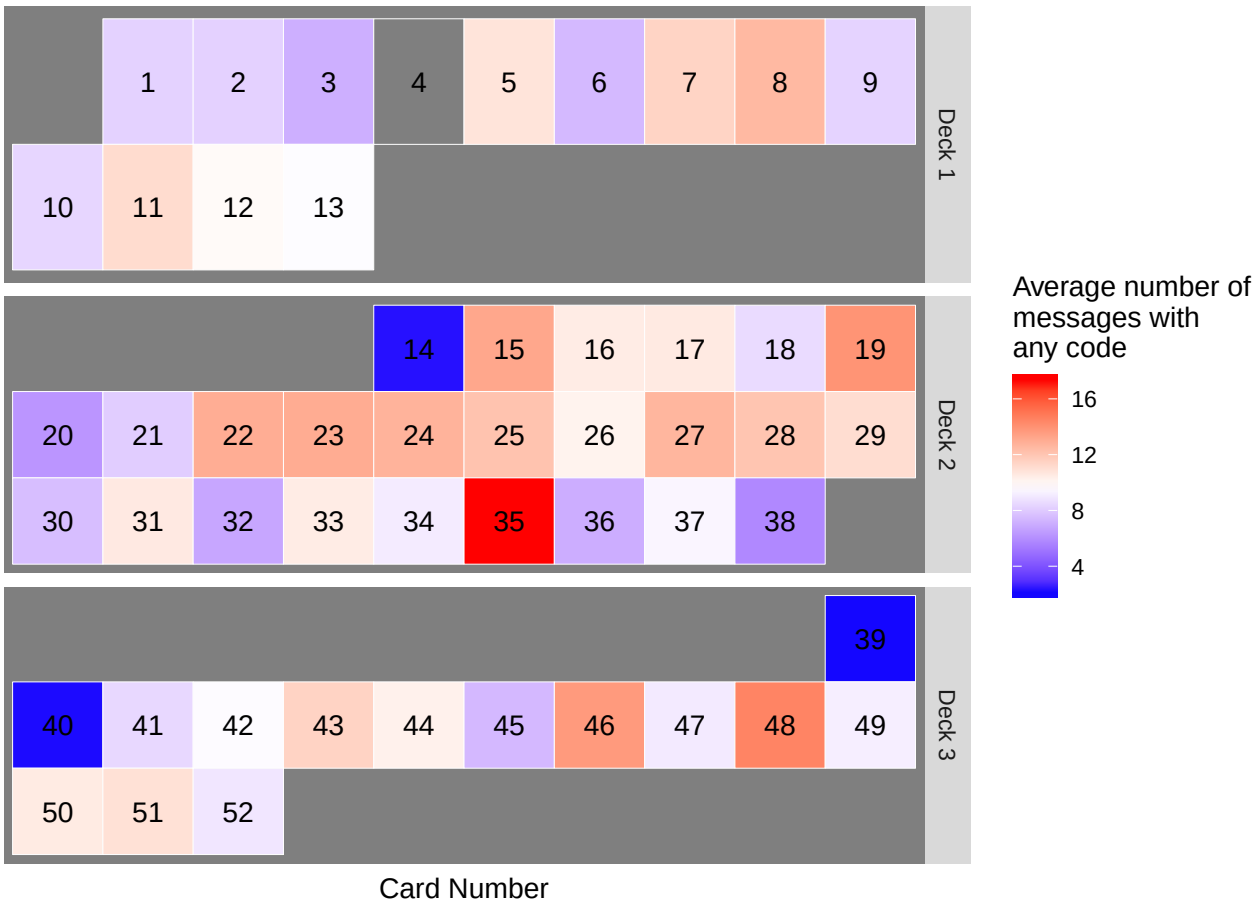

Figure 35: Average count of any category per card

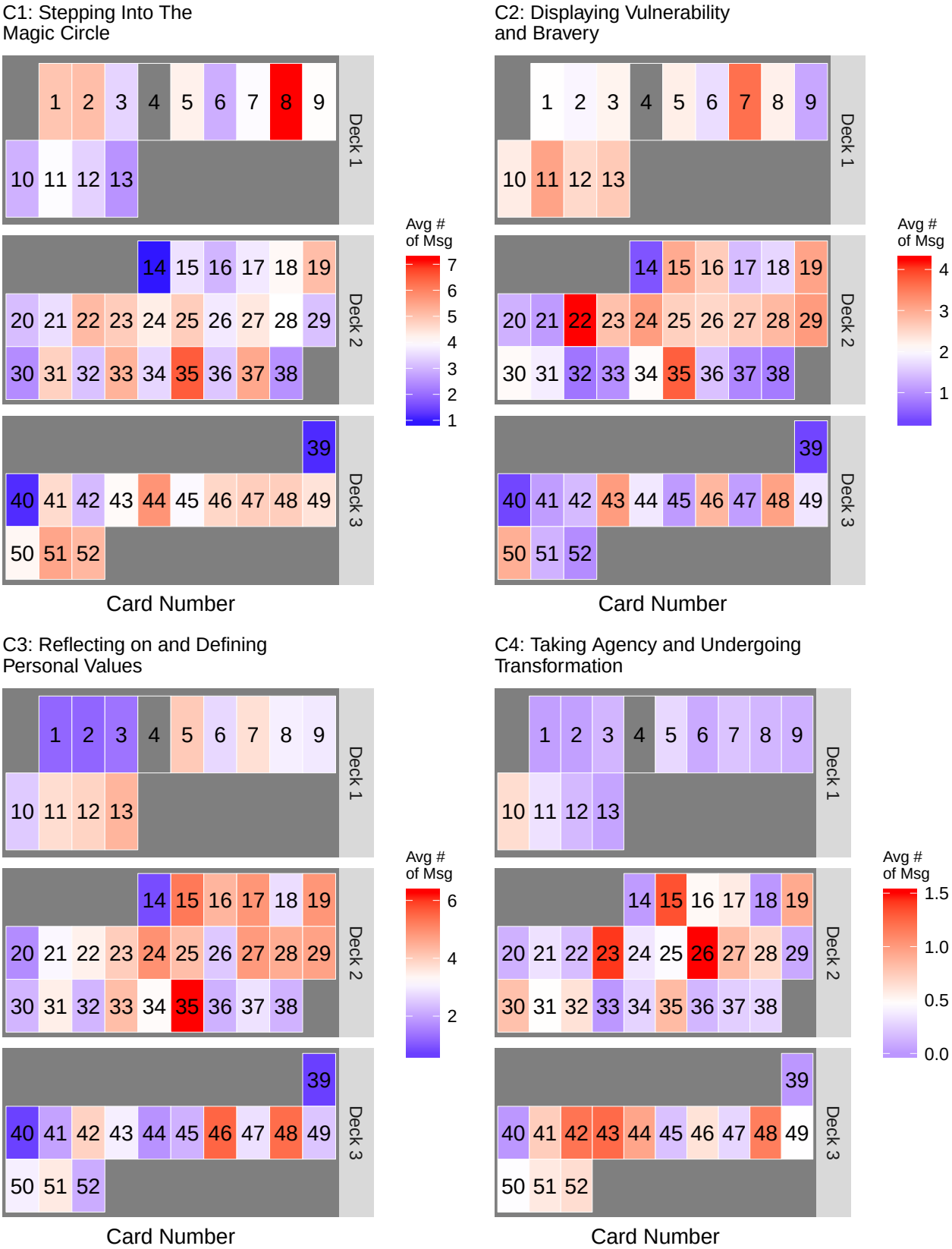

Figure 36: Average count of each category per card

Finally, we consider whether different activity types elicited different frequencies of player map coded messages. Figure 37 plots the average number of per-card messages in for cards under each category. While there is not too much variation, we do see that cards characterized as "embodied experience" exhibit far fewer Category 4 interactions compared to the other 3 types.

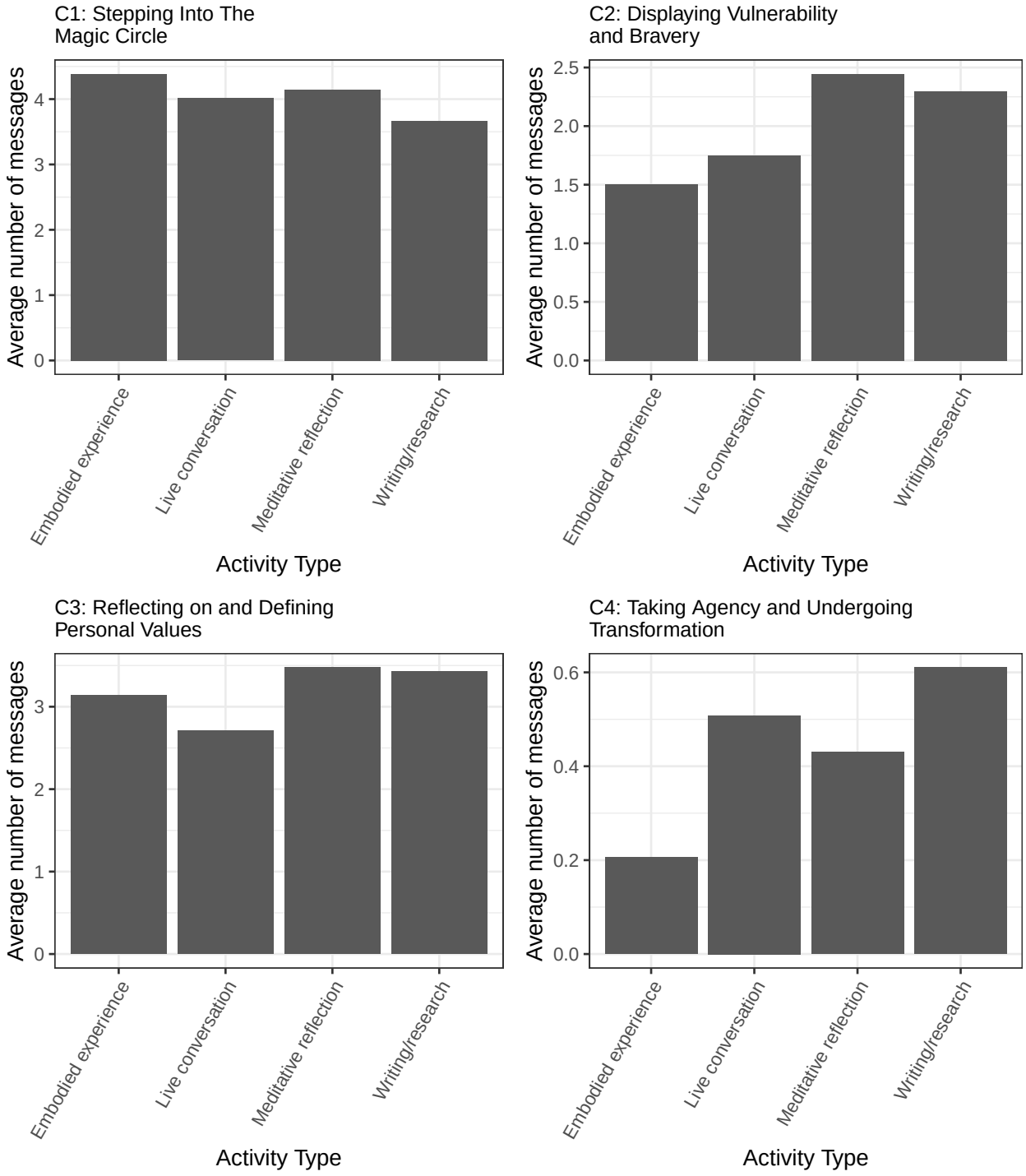

Figure 37: Player map frequencies by activity type