Virtual Pets and Virtual Selves as Exercise Motivation Tools

David B. Ramsay[1], Jenny Jin[2], Dr. Pattie Maes[1], and Dr. Roz Picard[1]

[1]MIT Media Lab, Cambridge, MA

[2]MIT Sloan School, Cambridge, MA

dramsay@mit.edu, jennyjin@mit.edu, pattie@media.mit.edu, roz@mit.edu

ABSTRACT

Virtual pet games- in which a user regularly cares for a small creature- have been effectively used as health behavior change agents with children. Research using virtual selves (avatars) have also provided promising results. In this paper, we outfitted two groups (N=20) of college-age students with fitness trackers that controlled either (1) the health and happiness of a virtual pet or (2) the weight and happiness of a virtual avatar, and compared the effects on their exercise patterns over the course of two weeks. We also measured participant narcissism, empathy, and self-image to compare psychographic data against exercise motivation and pet/avatar attachment. Differences in exercise behavior where not significant, however the human group had higher engagement and attachment to their avatar. The data suggests that empathetic people may grow more attached to pets, but with such a small sample this claim requires further verification.

Author Keywords

Avatar; Virtual Pet; Tamagotchi; Exercise Motivator; Behavior Change; Gamification

BACKGROUND

Tamagotchi was a hugely successful game in the 1990s with children, in which they were required to feed and care for a virtual pet on a keychain. The addictive nature of the game note only led to financial success, it also led to them being banned from several schools for distracting the kids.

This concept has been used recently to help motivate healthy behavior in children. A 2012 study by Byrne et. al. [1] showed that virtual pets can lead to twice as many healthy eating choices over a nine day trial with adolescents, and that having a negative feedback condition (i.e. a sad pet) was required to successfully promote these behaviors and the corresponding feeling of pet attachment. These results are very hopeful for health intervention design.

Commercial products have taken advantage of this technique. Products such as 'LeapBand' and 'Ibitz' are both wearable pedometers marketed towards children, in which they control and care for a virtual pet with their exercise. 'Walkamon' is a pedometer based iPhone application geared towards the general market in which little monsters grow as you meet your walking goals. The idea of a 'Tamagotchi' for gamified health change has a clear market, though product designs are not yet incorporating best practices from the research.

In a similar vein to virtual pets, there has been a lot of recent work to understand how avatars, or virtual representations of ourselves, may be used as a tool for behavior change as well. The work by Jesse Fox et. al. in particular stands out. In one paper on the subject, from 2011, he shows that realistic happy/sad renderings of a participant's future self had a significant impact on their retirement saving decisions [2]. His work from 2009 [3] showed compelling results, as participants watched realistic avatars of themselves in various conditions, including exercising, eating healthy/unhealthy foods, and gaining/losing weight. This series of studies included promising results for real-world behavior change.

Avatars have been a growing area of research in recent years, with numerous publications related to behavioral intervention. The results are promising in nearly every case [4]. Avatars have show positive impact as addiction counselors for all ages, as a clinical tool to help depressive young adults, and as a retail enhancement platform for seniors [5,6,7]. In fact, studies have shown that even in video games where behaviors are not meant to effect external behavior, there are clear impacts on self-conception and action in the real world [8].

Most of the avatar studies focus on immersion and realism as a crucial driver of vicarious reinforcement and self-modeling. In the case of the Fox et. al. work, they use real pictures of participants to create the virtual avatar, and use immersive 3D virtual reality devices during their studies. In the exercise related cases, having a negative feedback case was not shown to be important (as it was with the pets), perhaps because the driving mechanism is fundamentally different. It seems that identity, reinforcement, and consistency underlie the changes in behavior with an avatar when the outcomes are tightly coupled and well understood, instead of a desire to take care of a pet with which you empathize. We are interested in creating a real world, engaging intervention for health behavior. We want to understand how well the principles of a realistic, immersive 3D avatar translate to a scalable, Tamagotchi-style game– on a mobile phone with a more enjoyable, cartoon-version of the user. How does a cartoon vs. realistic avatar change the efficacy of the intervention, and how does it change the user engagement? Should you see

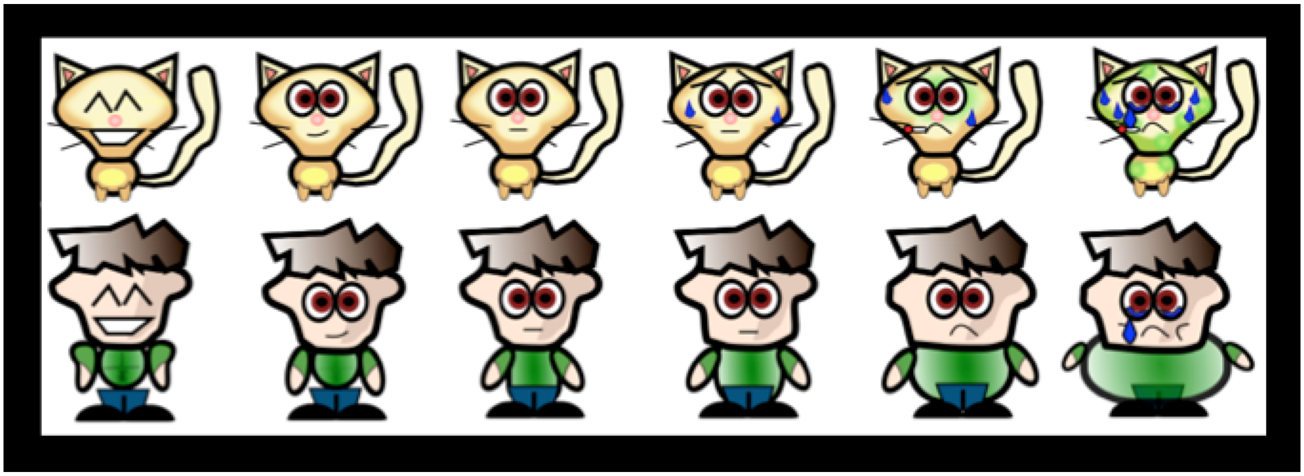

Figure 1. Virtual Pet and Virtual Self Physical/Emotional States (from 0 [left] to 5 [right])

your avatar working out all the time? Should your avatar have negative reinforcement (i.e. get fat)? How much does the emotional state of your avatar matter? Parallel questions exist for virtual-pet based interventions.

Before attempting to understand and optimize either the avatar or virtual pet design, we wanted to compare their efficacy. Both techniques show great promise, and both appeal to different underlying motivations. We believe we can develop either of them into an engaging, Tamagotchi-style game.

In this study, we compared two interventions– one based on a virtual pet, and one based on a virtual self. We collected user data on narcissism, self-image, and empathy in order to understand how psychographic profiles might effect the relative efficacy of both interventions. Finally, we measured changes in self-efficacy, attitudes toward exercise, and feelings of attachment to the pet/avatar in addition to the changes in exercise behavior.

STUDY DESIGN

Overview

In this study, we gave 21 college-age students a Fitbit wearable pedometer, which monitors steps, distance, and speed. We assigned 10 students a mobile application with a virtual pet, that had six states (from happy/healthy to sad/sick), and the other half were assigned an application in which they were asked to create a virtual self with had an analogous six states (from happy/fit to sad/obese). The state of their pet/avatar throughout the duration of the two-week study was tied to their Fitbit activity data- thus, the more they walked, the fitter, healthier, and happier their digital companion.

Upon entering the study, we provided participants with a survey covering previous exercise behavior, self-efficacy, narcissism, attitudes toward exercise, and empathy. Upon exit, we gave them another survey to re-measure their self-efficacy and exercise attitudes, as well as attachment to the pet/avatar. We also asked for qualitative feedback. All of the surveys were taken from peer-reviewed studies in this domain, and will be discussed below.

For the duration of the study, we sent each participant a daily text with a picture of their pet/avatar in the morning, along with a one line phrase. The state of their pet/avatar was based on an exponentially weighted average of their exercise behavior over the last three days, compared to a Gaussian fit of the participants previous exercise results. This should result in each participant seeing every state approximately equally, unless they continue to improve. The daily score we used was based on the speed and distance of a user's moderate to vigorous exercise, as well as the sedentary time of participants.

During the study, we collected data on exercise behavior from the Fitbits (average speed, distance, active time, sedentary time), we measured compliance (days where users failed to use and sync their Fitbits), and we looked at app engagement (how many times users checked on their pet/avatar outside of the daily text).

In our results section, we look at overall exercise behavior, compliance, engagement, final attachment to avatar/pet, and changes in self-efficacy between the two groups. The influence of affective traits (narcissism, empathy), as well as group assignment is considered. We hypothesized that narcissists might have stronger results with the avatar compared to the pet, while empathetic individuals might show the opposite effect (higher attachment to pets).

Technical Overview

Both the virtual pet and virtual avatar projects were built using Node and MongoDB for the backend, and jQuery Mobile and Javascript/HTML for the frontend. All of the animation/design was done using the SVG tool Inkscape. API calls to Fitbit, the OAUTH authentication, session and user management/security were all handled using various 'passport' libraries. Texting was accomplished using the twilio service. Bcrypt and zxcvbn were used for database/password encryption, Mongoose for database interaction, and Express/Jade for routing. Both sites were hosted on MIT servers using pm2 and nginx.

Table 1. Text Messages Sent with Pictures of Pet/Avatar Each Day

Fitness State *0=fit/healthy/happy* *5=fat/sick/sad*	Pet Message *example pet name is 'Snowball'*	Avatar Message *example user name is 'David'*
0	Snowball is deathly sick!! Only exercise can help!	Mini-David has really fallen off the wagon...
1	Snowball is quite ill, but a trip to the gym can cure it.	Mini-David is looking a bit large around the mid-section!
2	Snowball is starting to feel sick, don't forget to take care of your pet!	Mini-David is starting to show some chub- time to head back to the gym!
3	Snowball is doing alright- don't forget to plan your next gym trip!	Mini-David is looking alright- don't forget to plan your next gym trip!
4	Snowball is feeling healthy and happy, keep up the great work!	Mini-David is looking great, keep up the great work!
5	Snowball is the happiest!! Thanks for taking such great care!	Mini-David looks amazing!! Great work!

Surveys

We designed two surveys one for participants to take during onboarding and one upon exiting the study - with the goal to tease out some characteristics of participants personality traits.

In our onboarding survey, we asked participants ninety-four questions that focused on four assessment categories: narcissism, empathy, self-efficacy, and expected outcome predictions. We looked at extensive online literature on these particular personality scales that had already been designed for us to pull into our survey. We incorporated Hendin and Cheeks 12-question Hypersensitive Narcissism Scale. The HNSC test was built on questions from well-known assessment scales H. A. Murrays (1938) Narcissism Scale that correlated significantly with MMPI-based composite measure of covert narcissism [9]. We employed the 60-item Empathy Quotient test developed by Simon Baron-Cohen, which has been utilized to casually measure temperamental empathy for the general population [10].

In both the onboarding and exit surveys, we asked the same set of twenty-four questions that measured participants self-efficacy and outcome expectations of exercise. The goal was to look at whether participants confidence in their capabilities or predicted outcomes, which could affect their behavior, held steady over the course of the study. We incorporated the most well-known self-efficacy measurement created by Albert Bandura, and selected the fifteen questions that Bandura created which focused on regulation of exercise [11]. We also incorporated nine questions that measured outcome expectations for exercise developed by Barbara Resnick. Resnick argued that there was value in testing outcome expectations in predicting behavior apart from self-efficacy questions [12]. Thus we combined both sets of questions to round out a comprehensive assessment for looking at how participants beliefs on their capability and predicted outcomes may affect their behavior.

In our exit survey, we included a combination of quantitative and qualitative questions that measured participants attachment to their avatars pulled from a study by Sasha Byrne that looked at youths eating behavior through caring for mobile virtual pet avatar. Given the similarity of the mobile virtual avatars in assessing motivation levels of a desired behavior, we believed these questions were a good fit for our study in examining attachment scales. We incorporated the nine quantitative questions on a 5-point scale that probed attachment levels from Byrnes study. We also pulled in and adapted seven qualitative questions for participants to describe openendedly how they felt when, for example, their pet avatar went from being happy/healthy to sad/sick [1].

We relied on the scale values that each test assessment author created in calculating our participants survey scores. For all of the questions except for the self-efficacy section, the authors used a 5-point scale with a range from strongly disagree to strongly agree. The self-efficacy questions created by Banduras ranged from 0 (dont feel capable at all) to 100 (feel very capable) and we employed those numerical values across all participants.

Part of our study goals was to test whether certain personality traits had significant relationships with a certain avatar. We hypothesized that higher levels of empathy would have a positive relationship with attachment to pets, and that higher levels of narcissism might have a positive relationship with attachment to human avatars. Relying on these established surveys from psychology literature helped us to build a stronger assessment of the personality traits and attachment measurements that we hoped to gather from our participants.

In-Depth Review of Experiment

We recruited 21 students from the MIT Sloan Business School, ages 25-34, to participate in our experiment. They were randomly assigned to the two conditions (avatar and pet), totally 11 humans and 10 pet users. 11 participants were men and 10 were women, evenly split between the two groups.

Participants were recruited with a $25 gift card incentive, provided at the end of the experiment. We onboarded them in groups on a Monday, April 13th – giving them a Fitbit, having them review the COUHES consent form, completing our onboarding survey, and getting set up with both the Fitbit and our custom phone application.

Over the next five days, we verified that each participant had successfully signed up, customized their avatar or named their pet, and started using their Fitbit and syncing their data. The Fitbit servers had intermittent reliability issues during the two day period in which we were onboarding, and a couple of participants had issues using the Fitbit, so we followed up with each participant individually. During this time we started collecting baseline fitness data to use when deciding whether their pet/avatar should be healthy/fit or unhealthy/fat on proceeding days. Most of the participants were able to check on their avatar/pet and use their Fitbit as normal during this period, with a handful of delays due to technical issues (our most common issues were (1) unresponsive Fitbit servers, and (2) inadvertently capitalized usernames).

For both applications, the initial screen allows the user to login or sign up. Users are then immediately redirected to the Fitbit website (if they're creating an account for the first time) to authorize our app with the Fitbit API. From here, pet group participants are routed directly to the main screen (figure 2B), while the avatar group is sent to the avatar creation screen. The avatar group was asked to create an avatar that looked like them. There are 2 genders/body types, 16 skin tones, 18 eye colors, 18 hair colors, 16 shirt colors, 16 pants/skirt colors, 34 hair styles, and 8 extras (glasses, eyelashes, beards, mustaches, etc) to choose from (approx. 720 million possible combinations). A few examples of completed avatars are shown in Figure 2A. Pets and avatars are programmed to randomly blink (rapid and slow styles), breathe, and change their expression for a brief period to a similar expression (i.e. smile largely when grinning, or crying when sad) to bring them to life. Otherwise, the application is not interactive. All future sessions bring the user directly to the main screen, where they can see their avatar or pet. This screen has one button, which links to a 'settings' page (they can contact us, change their avatar, force their Fitbit to sync, or look at their fitness data). The only reason a participant would open the application is to view the changing appearance/state of their pet or avatar.

Figure 2. Example avatars (top) and the app's main screen (bottom)

From Saturday, April 18th to Sunday, May 3rd– a period of 15 days– we texted the participants every morning around 10am with a picture of their avatar/pet based on their exercise during the previous day. Each text included a phrase with it– these phrases are shown in Table 1. We had no interactions with any of the study participants during this time, and no one reported any technical issues. Syncing/texting data from the backend corroborated that no technical issues arose for participants during this period.

Participants were not contacted to sync if they failed to comply with the study. Generally speaking, users had 'background sync' selected, and were syncing dozens of times each day. On mornings when we hadn't yet received a sync notification from a user's Fitbit, we included the phrase *'- don't forget to sync your Fitbit today!'* at the end of their text. If it had been three days in a row with no sync data, we would text them a picture of their sad/unhealthy/fat state with the phrase *'You haven't synced your Fitbit in over 3 days! We have to assume the worst...'*, as shown in Figure 3.

Participants were not told to check the application, but given the option. We made it clear that we would text them during the study, so that opening the application was not necessary. We then monitored app usage to see if any users felt compelled to check on their avatar more than the once a day text. Any activity by a user in the application over a 20 minute sliding window was grouped together and categorized as a discrete event.

Besides counting the number of daily Fitbit syncs (compliance) and app engagement, we also cataloged daily Fitbit fitness information. Fitbit provides distance, steps, and several subcategories of step-related activity: very active, moderately active, fairly active, and sedentary. For each of these divisions, the Fitbit API provides duration and distance information at each level. What exactly these categories mean, unfortunately, is slightly obfuscated by Fitbit. However, anecdotally, a brisk walk can easily pass for 'very active' behavior.

To determine the appropriate current state for a participant's avatar/pet, we use 'very active' distance and speed. We also dis-incentivized sedentary time. Our final daily score was calculated in this way:

$$score = \frac{distance_{veryactive} * speed_{veryactive}}{time_{sedentary}}$$

$$= \frac{distance^2_{veryactive}}{time_{veryactive} * time_{sedentary}}$$

This rewards faster pacing, longer runs and harder work-outs, rather than extended periods of lighter activity (which the traditional pedometer 'step' model incentivizes). We told participants that their pet/avatar should respond to harder exercise. In practice, few participants varied their workout intensity in a significant way throughout the two weeks. This means that the variation in their daily score matched very closely to the trends in an individual's daily distance and steps, and inversely with sedentary time (which all vary together in a highly correlated way).

To take this daily score and end up with a representative state, a 'current score', using the weighted average of the current day and the past two days, is calculated:

$$score_{current} = score_n + 0.65 * score_{n-1} + 0.35 * score_{n-2}$$

where *n* is today, *n*-1 is yesterday, etc. If data from one of the days is not available (due to lack of compliance), it is simply omitted from the calculation. This helps to buffer dramatic changes in avatar/pet state, assuming the user has an exercise cadence of 2-3 days. The 'current score' is then compared to the mean and standard deviation of all prior personal daily scores to decide on the state of the pet/avatar, using the following logic (evaluated top to bottom):

Equation	Fitness State *0=fit/healthy/happy 5=fat/sick/sad*
(currentScore >(averageScore + stndDev))	0
(currentScore >(averageScore + (0.5 * stndDev)))	1
(currentScore >(averageScore))	2
(currentScore >(averageScore - (0.5 * stndDev)))	3
(currentScore >(averageScore - stndDev))	4
(currentScore ≤ (averageScore - stndDev))	5

This ensures each user (1) gets all types of feedback, in a Gaussian distribution, unless they consistently improve, and (2) makes sure that everyone has a baseline that is tailored to their fitness level and activity. No matter what your fitness level is coming into this system, you must improve and push above your normal athletic range in order to sustain a happy/healthy companion.

After the 15 days were completed, we followed up with each participant to complete an exit survey. We let them keep their Fitbits for 8-15 days after the end of the texting intervention, while recording Fitbit and app engagement data. As of yesterday, 10 days after the study was completed (and several Fitbits turned in), two human users were still checking on their avatars once every 2-3 days. All pet users stopped checking on their pet on the last day of the study.

Figure 3. Text/App when the user has failed to sync

RESULTS

Overview

Overall, we looked at exercise behaviors (distance, steps, active time, and sedentary time), app engagement (number of app openings), Fitbit syncs (compliance), attachment, change in self efficacy, and qualitative feedback as the outputs from this experiment. We compared this to our main test (avatar vs. pet groups), as well as looking at empathy and narcissism. Besides viewing overall pet vs. avatar for the entire group, we plotted data using only the users that reported a positive experience (since the negative experience of some users adversely effected their results). Finally, we examined data on an individual basis.

Quantitative Findings

Since we track every subscription update from Fitbit, we can easily see how frequently users are syncing their Fitbit, and thus how compliant they are with our study. Over the 20 days, participants synced between 90 and 1600 times each. 'Background syncing' is the reason why we see approximately 100 syncs a day for some participants.

More telling than the absolute sync total is the number of days during the study that a participant went without syncing at all. In the avatar group, eight of the eleven users had at least one day of lapsed use (averaging 3.9 lapsed days, six days being the most). The pet group had only five users lapse during the study (averaging 1.8 lapsed days). Figure 4A shows the compliance of a particularly noncompliant avatar user, complete with the novelty effect of the texting intervention beginning on April 18th.

Beyond compliance, we tracked how frequently users opened the application to check on their pet/avatar. This was explicitly optional for participants. Few users checked on their pet/avatar after the texting portion of the intervention began. However, one pet user checked on their pet an additional 19

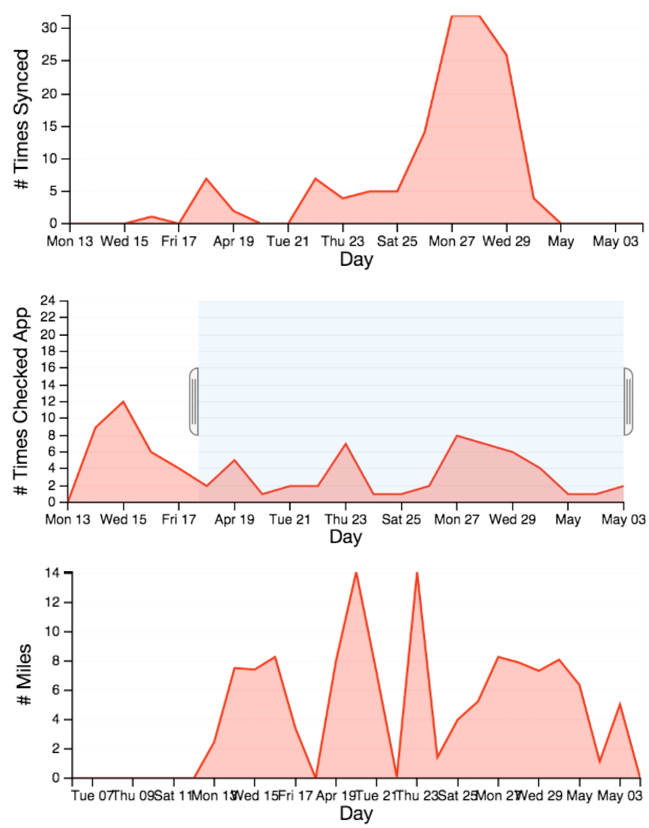

Figure 4. Examples of Compliance Data (# of Syncs), Engagement (# of App Checks), and Exercise Behavior (Distance in Miles) for Individual Participants

times during the 15 day intervention. On the avatar side, three users check on their avatars more than ten times throughout the two weeks. The single largest user engaged with their pet around 50 times in this period, averaging 3 times a day beyond the morning text (Figure 4B - the highlighted section corresponds to the texting intervention portion of the study).

We also analyzed steps, distance, very active time, and sedentary time to see if there were differences in exercise behavior between the groups. In general, these exercise metrics correlate well with each other for each individual, and we found extreme variability in fitness activity level from participant to participant. Over the full 20 days that users wore their Fitbits, distances from 22 miles to 140 miles per person were logged. Individual days ranged from 0 to 14 miles. On average, the avatar group logged 67.2 ± 32.4 miles, while the pet group logged 88.4 ± 27.0 miles. There is a pronounced novelty effect for several participants immediately upon receiving their Fitbit, and for about six participants we see a large novelty effect as well when the text-based intervention began (Figure 4C - the first day of the text intervention corresponds with the first large spike). Besides the wide spread in exercise behavior, there was also an interesting result of pet users both having more very active time *and* more sedentary time than their avatar counterparts (Figure 5C). Even on an individual level, fitness data did not show any strong trends or provide any clear insights. Hopefully with some of the post-study data we're currently collecting, we can get a better sense for how individual behavior might have changed in reactions to the intervention.

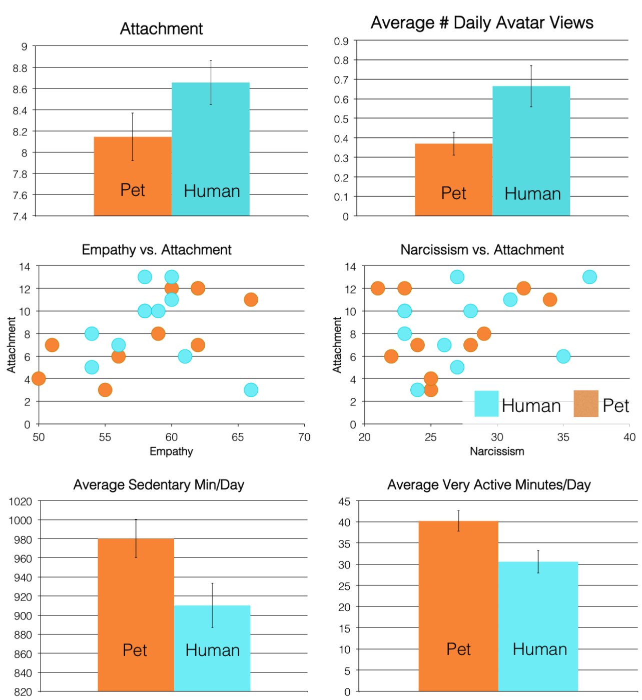

Figure 5. Average Attachment, Daily Engagement, and Exercise Results for the Two Groups

There is a significantly higher attachment to the virtual avatars over the virtual pets, mirroring the average application engagement data (Figure 5A). When plotted against empathy and narcissism, we see there is a slight trend in the pet group-more empathetic people tend to show higher attachment–which does not hold for the human group (Figure 5B). This is a weak correlation, but one that matches our hypothesis and is worth exploring further.

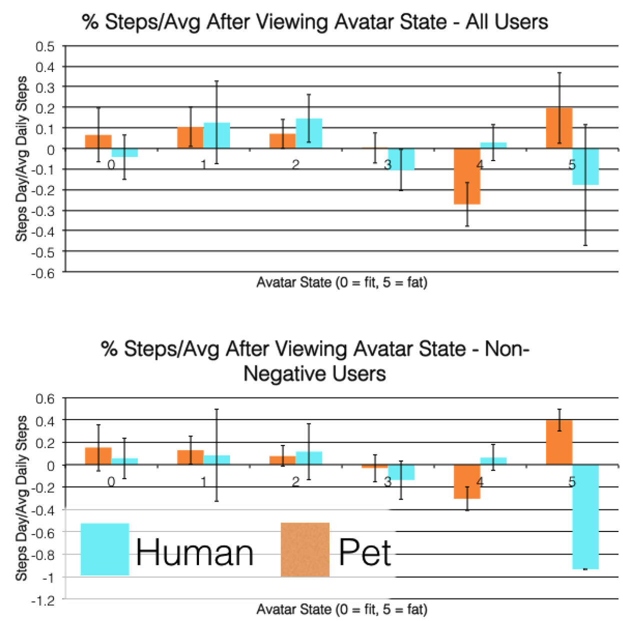

Figure 6. %Steps/Avg after Receiving a Certain Text State, for All Users and for Users that had a Positive Experience

The pet group showed a slightly higher positive change in self-efficacy after the intervention compared with before, but the results were very small (about 0.5% difference of the full self-efficacy scale on average) with a wide distribution. Even when looking at participants with the most positive qualitative experience, their self efficacy showed no clear trend.

We explored all of the above data (personality, exercise behavior, attachment, engagement) on an individual level for each of the participants. There were no clear trends beyond the parallels between engagement and attachment. Besides looking at aggregate data, we also looked for links based on the texts received. We checked active time, sedentary time, steps, and likelihood to check on your avatar after receiving a morning text in various states (from fit/healthy/happy to fat/sick/sad). We isolated the group with the strongest qualitative reporting (those that reported a good experience and emotional connection to their avatar), to see if they were more influenced by the texts, and/or demonstrated any other significant trends. We also explored individual text patterns throughout the two week study. There were no significant patterns. An example of this data is shown in Figure 6, in which the percentage of steps individuals took above or below their average was compared against the state of their virtual companion from that morning, both for the entire group and for participants reporting some qualitative emotional connection to their pet.

Qualitative Findings

Starting users off with the Fitbit seems to have produced a reasonable novelty effect for some users, which became a serious issue for a few of our participants. High motivation and achievement on the first few days can lead to repeated negative messaging throughout the two weeks, even when the participant feels like they're engaging in health behavior.

Additionally, we found that participants have a very strong internal notion of their health and fitness, as well as a strong notion of a healthy day for them. In some cases, participants reported that they had cycled or done yoga, and it didn't appear as though the Fitbit captured that information (because, of course, it doesn't). In other cases, the aforementioned novelty effect meant that users felt there was no causal link between their behaviors and the state of their avatar/pet. In both groups, users reported being frustrated, disillusioned, and apathetic.

Four of the ten pet users had a strong enough negative experience to completely detach. Their qualitative feedback included things like feeling *'indifferent, distant, disconnect'* from their pet, or that it was *'annoying, with no emotional connection'*. One participant just replied *'not accurate'* to all survey questions related to their emotional connection with the pet. Two of these four pet group users specifically mentioned not liking cats, and wishing they could've customized it or picked another animal in their feedback. Surprisingly, the user that checked on his cat the most, averaging 1.3 times a day above their morning text, fell into this category.

Six of the eleven human users reported similar sentiments, from *'indifferent'* and *'neutral'* feelings toward their avatar to *'apathy'*. Two of this group specifically stated that *'it didn't know what I ate'* or that *'it did not reflect the entirety of my physical activity and food intake'*, and one simply said *'everything I received from my avatar was negative'*.

The remaining six pet users indicated somewhat frustrated, though somewhat emotional connections to their pets. They described their pets as *'cute, but temperamental'* or said they felt a mix of negative (*'helpless'*) and positive (*'proud'*) emotions. There was one person in the avatar group who also fit into this category.

Finally, there were four people in the avatar group with decisively positive things to say about their avatars. They mentioned feeling *'responsible'* for and *'encouraged'* by their avatar, as well as *'caring and compassion'* towards it. Two people called their avatar *'an extension of who I was'* and *'part of me'*.

Three pet users indicated wanting to continue with the pet, one said maybe, and the remaining six said they would not like to continue. Four human users indicated wanting to continue with the avatar, two said maybe, and five stated they would not like to continue.

DISCUSSION

Overall, the quantitative fitness data was ambiguous, and didn't reflect any clear trends. This is not unexpected- these were small groups that included many participants of varied athletic background. The most clear and promising results are the attachment and engagement data, which clearly show increased efficacy for the avatar version of the application. Furthermore, the divisive nature of the human group- either highly noncompliant or highly engaged- stands in stark contrast to the virtual pet group. It suggests that the self-identity of the avatar strikes an emotional and impactful chord with the participants. The data supports the idea that it is a more engaging, impactful platform when built correctly.

The possibility that more empathetic people become measurably more attached to their virtual pets is also suggested by this data. This should not come as a surprise, though conclusive proof requires further investigation. It potentially does hint that despite the higher engagement we saw with the avatar group, a virtual pet could be very compelling to an empathetic individual. Furthermore, if we structure the interaction to increase the sense of empathy and bond (a cuter pet that requires more time investment and allows for customization, individuality, and choice), the pet design could work very well. Further work should be done to explore the possibilities of a customized virtual companion based on a user's personality.

There were clearly several issues with causality, and several best practicing moving forward come directly from these results. The qualitative data is particularly helpful, and hopeful, in providing direction for the next revisions of this work. This user study gave us valuable feedback about user expectations, mechanics, and connection, that we can address in future work.

FUTURE WORK

With backing from the Undergraduate Association, we are hoping to continue this work in a serious way over the course of the next year. We're planning to design for engagement, and will focus on building in easy to use data entry and an immersive, gamified structure. This will lead to a sustainable user base, who we can use to experiment with new features and ideas to rapidly optimize the platform for engagement and health outcomes.

We have the ability to quickly run a very similar second round user study, with the Fitbits from the first study, so we are considering collecting more data using the current platform after addressing a few of the main lessons learned (below).

It would be very compelling to specifically test for and correlate avatar attachment with exercise behavior results. Avatar attachment is quick and easy to measure and does not require a Fitbit, so verifying this logical assumption (that feeling emotionally connected to the motivational agent leads to more impactful behavioral outcomes) would make it easier to iterate and test several small assumptions about virtual pet design with high confidence.

Even without proving this link, it may still be prudent to run a few quick, simple studies using the modified CASS (attachment to pet) scale. Despite direct proof, attachment and engagement certainly point to an better user experience, and a stronger starting point for introducing behavior modification techniques than a less emotionally engaging design.

One example of a small study we might run would split participants into two groups after providing everyone with a standard virtual pet. The difference would be that one group gets to customize their pet, and one group cannot. After two weeks, we can assess their attachment with a short survey and interview, and draw some basic conclusions about the importance of customization. In a similar manner, we can evaluate levels of realism, the relative importance of health or happiness, and/or revisit this human and pet question with a much larger sample size (since Fitbits aren't required).

There are hundreds of unanswered questions about designing a Tamagotchi system that warrant exploration. It's possible to look at a small portion in this research context, but it can be a relatively slow process with a confluence of factors obfuscating clear results. It may be worthwhile to design experiments that probe more deeply into fundamental psychological motivations and processes that create success with these products. On the other hand, building a full system that can scale, and experimenting/iterating quickly on a reasonably sized initial user base could also yield powerful results. Using this model, it should be possible to optimize even the most subtle design decisions in a relatively short period of time.

LESSONS LEARNED

There were two main things we would've changed about our study design in hindsight. The first is the lack of customization for the pet version of the application. While there were clearly stronger bonds to the avatar, the lack of customization makes it hard to argue that these bonds are solely due to the human vs. pet condition, as oppose to investing time through a customization process, having an avatar that you know is uniquely your own, and/or the positive emotions that may come with some form of virtual companion selection instead of simply being given one.

The second issue is tied to the small group sizes and large variability in participant exercise behavior. There were people in each group who do very little formal exercise, as well as people who ran 10-14 miles in a day during our study. While these differences will average out with a large enough sample, our small sample size (constrained by the large investment of providing a Fitbit to each participant) makes it nearly impossible to get a large enough group to count on averaging out these effects. We originally designed the study to target people with similar exercise habits (minimal/none), and thus didn't build in a baseline week to track baseline exercise behavior. Due to time pressure, we ended up recruiting a varied group. We believe comparing individuals to their own baseline behavior would've been the best way to extract some meaning from this data. That said we are going to be looking at post study Fitbit data relative to data collected during the intervention to see if there are any obvious trends, after we finish the study next week.

The use of highly motivated, and highly stressed, business school graduate students also provides a very specific context to this experiment. The underlying psychology of this group is unique– it is clear that these students have high self-efficacy, a robust health identity, and pre-formed exercise habits. Does presenting someone with a strong internal self-image an antithetical external image (like a fat version of themselves) make him or her work harder? Our results seem to suggest it simply desensitizes them to the intervention. Does presenting them positive reinforcement motivate them? In many cases, it seems that they *already* reap the benefits of a strong self-image. Do these findings translate to other groups or individuals? We believe the answer is no.

Generally, the issues preventing Sloan students from exercising are scheduling related, and many of the users in this study were already motivated, aware, and practicing a routine. Someone in this position will be much less likely to respond to stimulus of this sort than a lower motivated individual with a weaker self-image. These individuals– who are not limited at all by time, but instead by psychological factors– are much more likely to show promising results, and likely will respond completely differently to an externalized identity if their own is weak or demotivating. BJ Fogg's idea of 'success momentum' through tiny psychological nudges applies much more to this population. As we saw throughout the behavior change course, slight changes in context can produce completely different behavioral results. We are more excited to solve this problem for infrequent gym-goers than highly successful business students, and we're not convinced the results will translate.

Finally, the causal link from health to avatar state is a very important thing to address. We saw participants instantly shut off from their companion when they felt that the link had broken down. There are two places where this can happen: from the user's behavior to what the Fitbit is able to capture, as well as the mapping from Fitbit data to the final avatar image. I was extremely surprised by the high expectations a non-technical audience had for the Fitbit tracker. Several participants seemed to expect all of their exercise to be captured (yoga and cycling included), and even food intake to be considered, despite the fact that the Fitbit is merely a pedometer. Making sure the participants believe this link is earnestly reflected by the application should be a top priority in all future designs. We believe this means (1) thoroughly explaining and setting expectations for what types of activities will be tracked, (2) providing easy/quick ways to enter other forms of fitness/health data to the application, and (3) letting the user have some say in the goals and expectations, so the system can adapt to the right level of 'harshness.' Ultimately, driving engagement is our top priority. You can't nudge anyone to change their behavior if they are turned off from your app and don't want to use it.

CONCLUSION

This study represents a first step in an extremely promising and highly under-addressed area for health behavior change. Like other novel research, these results beg many more questions than they answer. Overall, the avatar group showed higher attachment to their companion. How that ultimately manifests itself in behavior modification, however, will require a larger data set and a longer study.

In the short term, the qualitative feedback and design process have empowered our team to rapidly move forward with novel, compelling, and informed intervention design in the near future. It has also taught us how to structure successful user studies. We look forward to continuing this effort, armed with a strong foundation in behavior change design.

REFERENCES

[1] Sahara Byrne, Geri Gay, J. P. Pollack, Amy Gonzales, Daniela Retelny, Theodore Lee & Brian Wansink (2012) Caring for Mobile Phone-Based Virtual Pets can Influence Youth Eating Behaviors, Journal of Children and Media, 6:1, 83-99, DOI: 10.1080/17482798.2011.633410

[2] Hal E. Hershfield, Daniel G. Goldstein, William F. Sharpe, Jesse Fox, Leo Yeykelis, Laura L. Carstensen, Jeremy N. Bailenson (2011) Increasing Saving Behavior Through Age-Progressed Renderings of the Future Self. Journal of Marketing Research: November 2011, Vol. 48, No. SPL, pp. S23-S37.

[3] Fox, Jesse, and Jeremy N. Bailenson. "Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors." Media Psychology 12.1 (2009): 1-25.

[4] Murray, Tylar, et al. "Avatar interfaces for biobehavioral feedback." Design, User Experience, and Usability. Health, Learning, Playing, Cultural, and Cross-Cultural User Experience. Springer Berlin Heidelberg, 2013. 424-434.

[5] An, Lawrence C., et al. "A randomized trial of an avatar-hosted multiple behavior change intervention for young adult smokers." JNCI Monographs 2013.47 (2013): 209-215.

[6] Lisetti, Christine L., et al. "Building an On-Demand Avatar-Based Health Intervention for Behavior Change." FLAIRS Conference. 2012.

[7] Wiederhold, Brenda K. "Avatars: changing behavior for better or for worse?." Cyberpsychology, Behavior, and Social Networking 16.5 (2013): 319-320.

[8] Yoon, Gunwoo, and Patrick T. Vargas. "Know Thy Avatar The Unintended Effect of Virtual-Self Representation on Behavior." Psychological science (2014): 0956797613519271.

[9] Hendin, Holly M. and Jonathan M. Cheek. Assessing Hypersensitive Narcissism: A Reexamination of Murrays Narcism Scale. Journal of Research in Personality vol. issue 31 (1997): 588599. Web.

[10] Baron-Cohne, Simon and Sally Wheelwright. The Empathy Quotient: An Investigation of Adults with Asperger Syndrome or High Functioning Autism, and Normal Sex Differences. Journal of Autism and Developmental Disorders vol. 34, No. 2 (2004). Web.

[11] Bandura, Albert. Chapter 14: Guide for Constructing Self-Efficacy Scales. Self-Efficacy: The Measure of Control. Macmillan, 1997. 307-335. Print.

[12] Resnick, Barbara, Sheryl Itkin Zimmerman, Denise Orwig, Anne-Linda Furstenberg, and Jay Magaziner. Outcome Expectations for Exercise Scale: Utility and Psychometrics. Journal of Gerontology vol. 55B, No. 6 (2000): S352S356. Web.