Designing and Evaluating Carrot, a Persuasive System for Improving Good-to-Bad Behavior Ratios

Thomas Gill
University of South Florida
thomasgill@usf.edu

Abstract

This paper details the design process and evolution of a mobile app called Carrot, designed to help people improve their good-to-bad behavior ratios. The app draws on theory on persuasive systems, decision fatigue, gamification, open loops, and reinforcement. The design process is based around the elaborated action design research framework, and comprises cycles for diagnosis, design, implementation, and evolution. We also outline a plan for the future evaluation of the artifact. The contributions of this paper include the novel construct of the good-to-bad behavior ratio, the design of a system to improve good-to-bad behavior ratios, and the implementation of a prototype app which implements said design.

Keywords: Decision Fatigue, Open Loops, Gamification, Persuasive Systems, Elaborated Action Design Research

1. Introduction

These days people are expected to make more decisions and attend to more tasks and projects than ever before. Stephen Covey (2020) posits that people often focus on urgent tasks, both important and unimportant, at the expense of tasks and projects that are important but not urgent. Important projects that are unfinished and often completely unaddressed result in open loops in a person's mind which have the potential to increase stress and reduce focus (Allen, 2015; Heylighen & Vidal, 2008; Masicampo & Baumeister, 2011a).

When people do not have a clear objective and are worried about unfinished projects, it can get in the way of achieving a flow state, characterized by intense focus, confidence in one's performance and the feeling that the task at hand is intrinsically rewarding (Nakamura & Csikszentmihalyi, 2014). However, the stress caused by unfinished projects can be mitigated if an individual has a definite plan for attending to said projects in the future (Masicampo & Baumeister, 2011b).

To make matters more complicated, many individuals not only have projects they want to engage in more, but also have activities they could like to cut back on such as unhealthy eating, television watching, and social media usage. People not only have good behaviors they want to cultivate, but also bad behaviors they would like to mitigate. Making many decisions regarding what to do next or which activities to refrain from throughout the day can result in decision fatigue, which is liable to weaken one's willpower (Baumeister & Tierney, 2012).

In order to help people develop the plans and habits necessary to attending to their myriad tasks and projects in the important but not urgent quadrant of the Covey matrix, I developed a prototype for a mobile app called Carrot. This prototype revolved around the novel construct of the good-to-bad behavior ratio, which will be addressed later in the paper. The app was designed to provide users with a fun, gamified way of keeping track of all of their projects and desired tasks, to monitor their progress on tasks on a day to day basis, and to help users gradually dial back their undesired habits or behaviors over time.

The prototype was accepted to and presented at the Workshop on Information Technologies and Systems (WITS) in 2022, and the feedback was positive overall. The goal of this paper is to document in a more rigorous way the design cycles involved in implementing the 2022 version of the carrot, and in the archetype's evolution. These design cycles utilize the elaborated action design research framework outlined by Mullarkey and Hevner (2019). Using this framework to build a design artifact with the purpose of helping people cultivate desired behaviors while mitigating undesired behaviors yields three design contributions.

1. The novel construct of the good-to-bad behavior ratio.
2. Design models for an artifact which helps to improve the good-to-bad behavior ratios of users.
3. The implementation of the mobile app Carrot, which uses the aforementioned design to address the

URI: https://hdl.handle.net/10125/106529
978-0-9981331-7-1

HICSS

issues of decision fatigue, unfinished projects, and the good-to-bad behavior ratio.

My paper is organized as follows. In section 2, I outline the kernel theory from literature on persuasive systems and decision sciences which informed the design and implementation of my artifact, including already existing applications which address some of the same issues. Section 3 describes the elaborated action design research cycles involved in developing and refining the prototype application. Section 4 describes the initial diagnosis cycle, yielding the novel construct artifact of the good-to-bad behavior ratio. Section 5 describes the subsequent design cycle in which I produced a design model for the instantiation of an artifact which has the potential to improve the good-to-bad behavior ratios of users. Section 6 describes the implementation cycle and displays the instantiation of the artifact in the form of a prototype for a mobile application. Section 7 describes the prospective means of evaluating the artifact. Section 8 describes the implications and contributions of Carrot's design and implementation. Section 9 gives a summary of the research and describes future research plans.

2. Background

The first design cycle of the elaborated action design research (eADR) framework is the diagnosis cycle, which corresponds to the problem formulation stage of action design research (Sein et al., 2011; Mullarkey & Hevner, 2019). The problem space for Carrot's design included theory on habits, will power, gamification, and open loops. It also comprised previous artifacts addressing these issues, as well as best practices for the design of said artifacts. In sum, my design draws on both the descriptive theory of decision sciences, and the prescriptive theory of persuasive systems, and decision support systems. Figure 1 provides an outline of the kernel theory which influenced the design and implementation of Carrot.

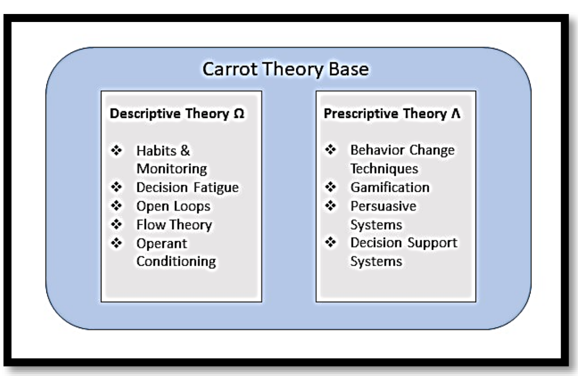

Figure 1. Kernel theory informing the design and implementation of Carrot app

Descriptive Theory

A habit consists of a cue, a routine, and a reward (Duhigg, 2012). The cue is a trigger from the environment which sets off the habit. The routine is the set of actions taken by an individual in response to the cue. The reward is positive reinforcement experienced by the individual as a result of the routine. People may not be conscious of habits, and may not be aware of the specific cues that trigger them, and the more a habit is repeated the stronger it becomes. The best way to change habits is to identify the environmental and emotional cues that trigger them and find a new routine with which to respond to said cues.

To resist the urge to respond to a cue with one's usual routine requires willpower, which gives people the strength to persevere or resist temptation (Baumeister & Tierney, 2012). Research has shown that decision fatigue is one of the key phenomena which depletes will power, as decision making effort depletes an individual's glucose which is required for the resistance of temptation (Wang & Dvorak, 2010). For instance, a previous experiment found that judges become less likely grant parole the later in the day it gets and the more parole decisions they've already made (Danziger, Levav, & Avnaim-Pesso, 2011). As decision fatigue sets in, people are more likely to conserve mental energy by choosing the default option, which in the case of the judges was to refuse parole. For this reason it has been proposed that many decisions could be improved simply by changing the default option, which is an example of nudging (Thaler & Sunstein, 2009).

From the research on will power, we can deduce that having to make many decisions regarding which projects to work on next can cause decision fatigue and reduce an individuals ability to resist temptation. An overabundance of unfinished projects can also result in open loops in the brain which reduce focus and cause stress (Masicampo & Baumeister, 2011a). This can hinder individuals from achieving a flow state characterized by stress free focus on a task (Nakamura & Csikszentmihalyi, 2014). Csikszentmihalyi (1993) outlines four requirements for achieving a flow state:

1. Each part of a task or activity must have a clear sub-goal.
2. There must be clear, unambiguous rules for the attainment of said sub-goals.
3. To further avoid ambiguity, an activity must provide constant feedback as the individual is engaged in it.
4. The difficulty of the task must be a good fit for the skill of the individual engaged in it.

The primary goals of my artifact, informed by decision sciences literature, is to promote better habits, reduce decision fatigue, facilitate a flow state, and to help motivate users to attend to tasks in the important but not urgent quadrant of the Covey Time-Management Matrix. We will look now at some prescriptive solutions which have previously addressed this problem space.

Prescriptive Theory

My goal in designing this artifact was to help people become more productive while cutting down on undesired behaviors. This goal naturally involves changing the behavior of individuals through the use of an information system. There already exist many interventions and applications designed to help users live more healthy and productive lives. These interventions make use of a wide variety of behavior change techniques which mitigate factors that can prevent behavior change or augment factors that can facilitate behavior change (Carey et al., 2019). A taxonomy of behavior change techniques is provided by Michie et al. (2013). Examples of existing behavior change interventions include a mobile health intervention system designed to mitigate the risk of coronary heart disease among women through the encouragement of exercise (Sengupta, Dutta, Beckie, & Chellappan, 2020), diet tracking applications such as LifeSum and MyFitnessPal (Ferrara et al., 2019), and a mobile app designed to deliver cognitive behavioral therapy for insomnia to its users (Horsch et al., 2017).

Persuasive systems are information systems designed to change the emotional or cognitive state of users (Torning & Oinas-Kukkonen, 2009). Previous research has outlined seven primary task support principles for persuasive systems: reduction, tunneling, tailoring, personalization, self-monitoring, simulation, and rehearsal (Oinas-Kukkonen & Harjumaa, 2009). Reduction entails the simplification of a complex task, process, or behavior. Tunneling entails guiding users through a process, persuading them to continue along the way. Tailoring entails fitting the functionality of the system to the needs of specific users. Personalization entails offering users personalized content. Self-monitoring entails that the system should help users monitor their own performance in the area that the system is meant to improve. Simulation entails that the system provide users with a cause-and-effect link between their actions and the results thereof. Rehearsal entails that the system provide users with a means of practicing or rehearsing their target behavior.

Decision support systems are information systems which aid users with decision problems (Liu et al., 2010). They have been applied to areas such as agriculture (Zhai et al., 2020), healthcare (Sutton et al., 2020), and transportation (Deveci et al., 2022). They often use statistical programming, simulations, or optimization models in order to inform users and improve their decision making in specified areas (Eom & Kim, 2006).

The goals of my design were two fold. First, I wanted to help people keep track of various projects and tasks to reduce stress. Second, I wanted to help people cut back on bad behaviors such as drinking alcohol or the consumption of junk food. At first glances these goals seem incompatible, as one involves the encouragement of certain activities, while the other involves the discouragement of certain activities. Examples of existing task management systems include the project management and collaboration tool Wrike (Rogers, 2014), the customizable work platform monday.com (Konrad, 2021), and Microsoft To Do, which allows users to record and categorize their various tasks. There are also applications which facilitate the reduction of undesired behaviors such as Nudge and MyFitnessPal which both allow users to record their daily food intake. These apps can help people lose weight as it has been shown that recording what one eats even once per week can result in dietary improvement (Hollis et al., 2008).

Aside from an applications usefulness, an application will be more effective if it is easy and even fun to use. One means of accomplishing this is through gamification, which has been defined as the enhancement of services through motivational gaming affordances (Hamari, Koivisto, & Sarsa, 2014). These affordances include rewards such as points, badges, achievements, and trophies (Hamari & Eranti, 2011). One of the key tenets of gamification is the matching of instrumental outcomes to experiential outcomes (Liu, Santhanam, & Webster, 2017).

In sum, there are many findings from decision sciences research, prescriptive research, and previous interventions that are relevant to the problem and solution spaces. I will outline the method I used to create and refine my design, in the following section.

3. Research Method

To develop my artifact I utilized the elaborated action design research (eADR) process model for the application of Action Design Research (Sein et al, 2011; Mullarkey & Hevner, 2019). The eADR process model was conceived by Mullarkey & Hevner (2019), to make the stages of Action Design Research, a design research method which involves an artifact that is shaped by an organizational context through a process of guided emergence (Sein et al., 2011), more explicit and clear to practitioners. The five activities of eADR are Problem Formulation & Planning, Artifact Creation, Evaluation, Reflection, and Learning. These activities and the underlying design principles are displayed in Figure 2a.

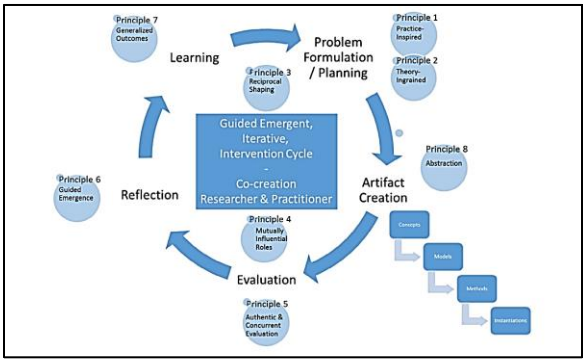

Figure 2a. The elaborated action design research (eADR) cycle (Mullarkey & Hevner, 2019, p. 8)

There are four stages of eADR (see Figure 2b), each of which represents a cycle containing the five eADR activities. The stages are as follows:

- Diagnosis: Involves an understanding of the relevant kernel theory and application domain.
- Design: Involves the conceptualization of an artifact's design
- Implementation: Artifact is instantiated in the form of a system, algorithm, process, database, etc.
- Evolution: An artifact evolves in tune with the requirements of a changing environment and problem space.

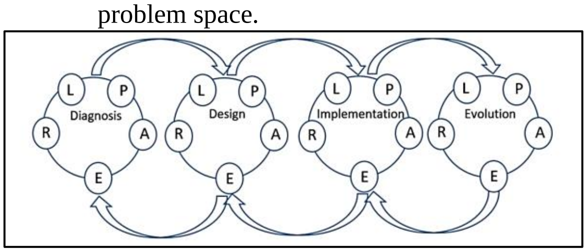

Figure 2b. The four ADR stages (Mullarkey & Hevner, 2019, p. 9)

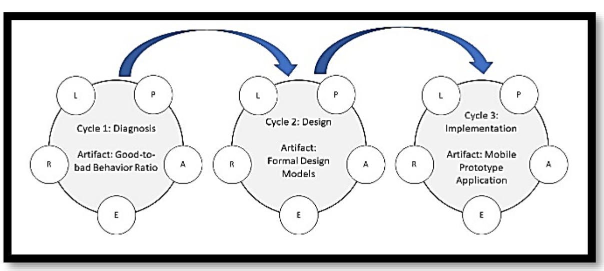

Figure 3. eADR cycles for Carrot

Elaborated Action Design Research is flexible in that researchers and practitioners can perform design cycles in the order they see fit. In designing my artifact, I went through three eADR cycles: a diagnosis cycle, then a design cycle, then an implementation cycle. Each cycle resulted in its own contribution. The contribution of the diagnostic cycle is a novel construct. The contribution of the design cycle is a design model. The contribution of the implementation cycle is a prototype mobile application. These cycles are outlined in figure 3. Each cycle yields a contribution at a different level of artifact abstraction as outlined in Gregor & Hevner, 2013. The three cycles of my artifact design are outlined in the following three sections.

4. Diagnostic Cycle

There is a parallel between the overarching goals of the mobile health apps referred to in section 2 and the goals of my design, viz., in order to facilitate healthy behavior, an intervention is going to have nudge users into performing more of some activities such as sleeping and exercising and less of others such as drinking alcohol or consuming junk food. The key difference is that mobile health apps are often informed by human experts such as health coaches, so that they are informing the users regarding which activities are desired and which are undesired from a health perspective, whereas with my app, I want to allow the users to be able to decide which activities they would benefit from performing more often, and which activities they would benefit from limiting or performing less often. I also do not want these activities to be limited to a purely health and fitness context.

To address the two goals of increased progress on tasks and mitigation of undesired behaviors, I have devised a novel construct, unaddressed in similar implementations, viz., the good-to-bad behavior ratio. This construct is useful, because in addition to increasing the frequency of desired behaviors such as reading, exercise, etc., people also often have behaviors that they would like to reduce such as overeating. The difficulty posed by this construct is that it is ambiguous and difficult to quantify. It could mean different things to different people. It could mean the ratio of dollars spent to dollars earned or the ratio of hours of working out to cheat meals consumed or the ratio of time spent reading to time spent watching television. This begs the question: how do I help people improve their good-to-bad behavior ratios, when said ratios are so difficult to define specifically and to quantify?

The answer is surprisingly simple. Although good-to-bad behavior ratios are difficult to quantify, it is relatively easy to quantify the improvement of said ratios, once a baseline has been established by a user. I would first ask an individual to make a list of good behaviors that he or she would like to cultivate as well as a list of bad behaviors that he or she would like to cut back on. As a proof of concept evaluation of the construct, I created a scenario which I went on to depict in a C# algorithm. Let's say we have an individual named Steve who would like to get more exercise, read more, and drink less soda. In the good behavior list Steve might include two good behaviors:

1. Exercise for one hour
2. Read 20 pages

In this case, list of bad behaviors would be limited to one item:

1. Drink one can of soda

Once these lists are compiled, Steve would define a starting ratio. If the ratio is one-to-one between good behaviors and bad behaviors, then for every can of soda Steve drinks, Steve will also have to exercise for one hour and read 20 pages. At this point, good-to-bad behavior ratio improvement is easy to quantify. If Steve wants to improve his good-to-bad behavior ratio by 50%, he will simply have to start exercising for one and a half hours and reading 30 pages for every can of soda he consumes. The result of the algorithm's execution is displayed in figure 4. The code of the algorithm can be provided by the author upon request.

This construct has the potential to give users a simple way of limiting the frequency of bad behaviors while increasing the frequency of good behaviors. If the bad behavior is something the user enjoys engaging in, it could also serve as a reward for the good behavior. After finding an index for improvement that would satisfy both the goal of increased frequency of desired behaviors and the goal of decreased frequency of undesired or unhealthy behaviors, the next step was to create a design which would facilitate the improvement of a user's good-to-bad behavior ratio.

```
1 hour Work out : Good
20 pages Reading : Good
1 bottle soda : Bad
Current Good to Bad Behavior Ratio: 1

1.5 hour Work out : Good
30 pages Reading : Good
1 bottle soda : Bad
Current Good to Bad Behavior Ratio: 1.5
```

Figure 4. Results of proof of concept algorithmic evaluation of the good-to-bad behavior ratio

5. Design Cycle

In order to form an initial good-to-bad behavior ratio, a user must first decide which behaviors to focus on. Each behavior must be accompanied by a specific unit. The unit can be time based such as one hour of working out, or it can be based on the completion of a task such as 20 pages of reading. The user must also assign frequencies to each activity, in the form of probabilities. Let's take, for example, a hypothetical user who is trying to cut back on eating out at restaurants, and is also trying to work out, read more, and meditate more. Said hypothetical user vies to work out for one hour, read 50 pages, and meditate twice for every restaurant meal consumed. Table 1 gives a behavior-frequency list for this hypothetical user.

Table 1. Behavior Frequency List for Hypothetical User

Activity	Amount	Frequency	Task Reward?
Workout	30 minutes	0.2	Task
Reading	10 pages	0.5	Task
Meditation	1 session	0.2	Task
Meal at Restaurant	1 meal	0.1	Reward

A user may have one of two goals with regard to a current good-to-bad behavior ratio. Those goals would be either maintenance or improvement. We will start by examining the maintenance option. We could have a user proceed through a check list of tasks to complete before being given a reward. However, this could be prone to cause decision fatigue, and may be difficult to navigate as the behavior list gets more complicated. The other option is a random approach, wherein once a task is completed, the system generates another item from the behavior list randomly using the specified frequencies. This relieves the user of deciding which task to complete next, or which order to complete tasks in. It also rewards the user for completing tasks on an intermittent reinforcement schedule which is effective at modifying behavior (Miltenberger, 2016).

A design model for a system which helps users maintain a specified good-to-bad behavior ratio is presented by the Stochastic Petri Net in Figure 5a. In this figure, P1 represents the number of tokens a user has available, P2 represents the number of pending tasks a user needs to complete, P3 represents the number of rewards available for consumption, and P4 represents tokens that have just been used to spin the wheel and is awaiting the results of the app (i.e. whether a reward has been earned or a task has been added to the pending tasks queue.) T1 represents a user's decision to use a token to spin the wheel, T2 represents the completion of a task, and T3 represents the consumption of a reward. T4 represents a token that has been transformed into a task and has a firing rate of λ. T5 represents a token that has been transformed into a reward and has a firing rate of μ.

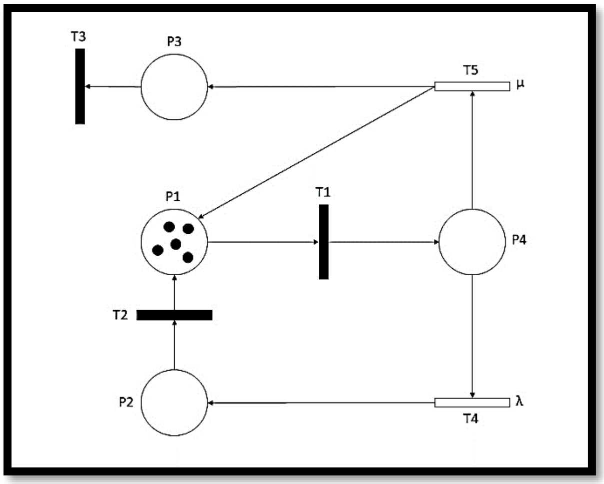

Figure 5a. Stochastic petri net for maintaining good-to-bad behavior ratio

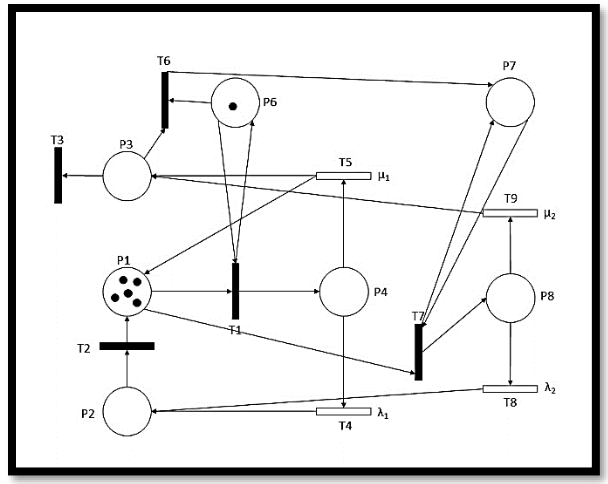

Figure 5b. Stochastic petri net for improving good-to-bad behavior ratio

With a simple extension of the petri net in figure 5a, we can allow for not just the maintenance of a good-to-bad behavior ratio, but an improvement of said ratio, represented formally by changes in firing rates λ and μ. In order to motivate the user to improve a behavior ratio, I associate progressively more desirable good-to-bad behavior ratios with the motivational gaming affordance of levels, which can be used as an indication of progression (Legaki et al., 2020).

The user to moves up a level simply by forgoing a reward, an indication that the user is ready for an improved good-to-bad behavior ratio to be enforced by the system. The Petri Net for this updated design is shown in Figure 5b. P6 and P7 keep track of a user's current level. If there is a token in P6, the user is at level one. If there is a token in P7, the user is at level two. T6 represents a levelling up, wherein the user decides to forgo a reward and P6's token is consumed and a new token a placed in P7. P8 represents tokens used to spin the wheel awaiting results for level 2. T7 represents the usage of a token at level 2. T8 represents a token being transformed into a task at level 2. T9 represents a token being transformed into a reward at level 2. The new Petri Net has four firing rates instead of two: λ1, λ2, μ1, and μ2. λ1 and μ1 represent the firing rates of level one for the generation of tasks and rewards respectively. Once a user transitions to level 2 by forgoing a reward, these firing rates are replaced by level 2 firing rates λ2 and μ2. λ2 is smaller than λ1 and μ2 is larger than μ1, resulting in more tasks completed on average per every reward received. By extrapolating this to several levels, we give users a way of improving their good-to-bad behavior ratios over time without having to consciously keep track of them.

The design in figure 5b serves as an abstraction for a system which could address both the issue of the desire to increase the frequency of certain behaviors and the issue of the reduction of bad habits and undesired behaviors.

6. Implementation Cycle

I instantiated the good-to-bad behavior ratio improvement design model in the previous section through Carrot, an android app which allows users to compile a list of tasks as well as a list of rewards. Users also choose a frequency for each task or reward in the form of a probability. Once the list has been compiled, a user starts with five tokens that they can use to spin a wheel. Each spin of the wheel generates either a task or a reward on the user's list in accordance with the user's specified frequencies. If a task is generated, a token is consumed. The user can earn more tokens by completing pending tasks. When a reward is generated, a token is not consumed. A generated reward is put in a pending rewards queue. While there is a limit to pending tasks, there is no limit to pending rewards. Each time a task is completed, or a reward is generated, the user earns a point. These points are used to keep track of a user's daily productivity index.

Whenever a reward is generated, the user has the option of either keeping said reward or forgoing it. If the user chooses to forgo the reward, the user levels up. A movement to the next level results in a shift in the user's good-to-bad behavior ratio. The frequency of good habits increases, and the frequency of bad habits decreases. Points and levels are gamification elements which have been shown to have a positive effect on performance (Mekler et al., 2013). Figure 6 contains eight screens displaying the app's functionality.

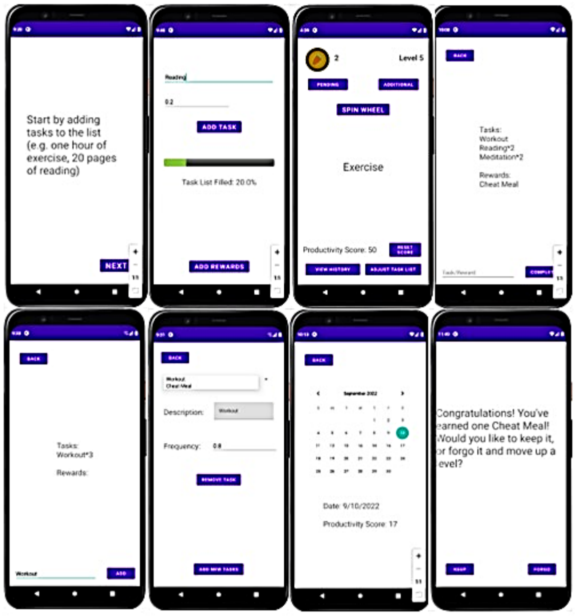

Figure 6. Screens from Carrot app

As I was the only person involved in the programming, the app is likely a suboptimal implementation of the design model presented in figure 5b. However, the implementation is still useful for evaluation purposes. The next section presents a detailed plan for evaluating the Carrot app and by extension Carrot's design model.

7. Intervention & Summative Evaluation

Elaborated Action Design Research calls for an evaluation to be performed during each of its stages. During the diagnosis stage, I performed a proof-of-concept evaluation by way of the algorithm presented in figure 4. During the design stage, I evaluated the design models by running a simulation of the petri net in figure 5b. The code and results of this simulation can be provided by the author upon request. During the implementation stage, I performed unit testing on the relevant java code as well as integration testing on the Kotlin code which integrates java code with the xml screens. I also performed scenario tests where I entered lists of tasks and rewards and completed them over the course of a day. I performed these tests both on an android virtual device emulator and on my own android device.

Venable, Pries-Heje, & Baskerville (2016) outline a framework for evaluating design artifacts composed of two dimensions: ex post vs. ex ante and naturalistic vs. artificial. The proposed future evaluation of Carrot's design and implementation comprises three phases, each addressing different design traits from Prat et al., 2015. In phase 1, I examine the artifact through the lens of theory from the Information Systems and Decision Sciences literature. In phase 2, the artifact is tested in an artificial environment to assess the degree to which it is accepted by users. User acceptance is predicated on the perceived usefulness of a technology and its perceived ease of use (Davis, 1989). Phase 2 will evaluate the app on both of these metrics. After phase 2, the artifact will be modified based on user feedback.

In phase 3, the artifact's usefulness will be evaluated through a longitudinal field study wherein a group of users are given the app to use over a period of a few months, detailing their progress in their projects, tasks, and goals along the way in a series of surveys and semi-structured interviews. Phase 1 corresponds to Hevner's (2007) rigor cycle, while Phases 2 and 3 correspond to the relevance cycle. These evaluation phases are displayed in figure 7.

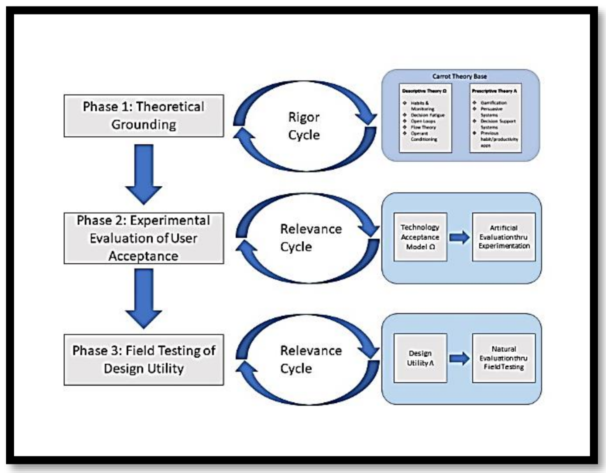

Figure 7. Evaluation of Carrot app

Table 2 outlines the ways in which Carrot's design and implementation address the primary tasks support principles and dialog principles of persuasive systems. Table 3 outlines the ways Carrot implements decision science principles to improve good-to-bad behavior ratios, mitigate the presence of open loops in the mind, and enable users to attain a state of flow.

Table 2. Persuasive system principle in Carrot

Primary Task Support	
Reduction	Carrot allows users to maintain and improve their good-to-bad behavior ratios without keeping track of their frequency of either. It also provides a productivity

	index to help them keep track of their daily productivity.
Tunneling	Carrot guides users through the process of using its system and improving good-to-bad behavior ratios. An increase in good behaviors and decrease in bad behaviors is simplified via the levelling up capability of the app.
Tailoring & Personalization	Carrot allows users to choose the behaviors they want to keep track of as well as their starting good-to-bad behavior ratio which they will later improve upon.
Self-Monitoring	Carrot allows users to effortlessly keep track of their good-to-bad behavior ratios and to improve them incrementally. It also provides users with a productivity index allowing them to estimate how productive they were on a given day.
Dialog Support	
Praise	Carrot congratulates users when they move up a level by forgoing a reward
Rewards	Carrot doles out rewards specified by the user intermittently. It also provides rewards in the form of points for completing tasks and level ups for forgoing rewards.

Table 3. Carrot's implementation of theories from the decision sciences

Operant Conditioning	Carrot provides users with intermittent reinforcement by way of rewards that users choose for themselves.
Decision Fatigue	Carrot chooses from a list of important tasks randomly to save the user from having to decide which task or project to address next. It also provides users with a bright line to help them abstain from undesired behaviors above a specified frequency.
Open Loops	By allowing users to compile a list of tasks that they know will be addressed eventually, Carrot reduces open loops caused by unfinished and unaddressed tasks or projects.
Monitoring	Carrot allows users to monitor the proportion of good behavior to bad behavior that they deem relevant, as well as keep track of their daily productivity in the important but not urgent quadrant.
Flow	Carrot keeps track of users' important projects and keeps users focused by feeding them one task at a time. It provides immediate feedback in the form of points and the ability move up levels and improve their good-to-bad behavior ratios.
Gamification	By allowing users to score points by completing tasks, and level up over time, the app is made to feel like a game. Its intermittent reinforcement schedule also represents an incorporation of game elements from slot machines.

Tables 2 and 3 show the theoretical effectiveness of Carrot based in the literature of decision sciences and persuasive systems. The experiments and field testing of evaluation phases 2 and 3 will help to assess artifact's relevance and improve the utility of the artifact.

8. Discussion

One of the contributions of this paper is the novel construct of the good-to-bad behavior ratio. Although this ratio is difficult to quantify in a general way, it is easy to quantify its improvement once the relevant habits have been chosen by an individual. Thus this construct provides a new goal for future persuasive and decision support systems to strive for, viz., to help people improve their ratios of good habits to bad habits. Another contribution is the design model presented in figure 5b which provides a road map to implementations for persuasive systems that can help users improve their good-to-bad behavior ratios, become more productive, and keep track of their unfinished tasks, goals, and projects in the important but not urgent quadrant, resulting in fewer open loops and greater flow. The third contribution is the implementation itself, on which we can begin evaluating the design's utility and fitness. Utility is defined by an artifacts usefulness in accomplishing what it sets out to accomplish, while fitness is the measure of an artifact's ability to evolve and replicate over time (Gill & Hevner, 2013).

Due to the random aspect of the artifact, it is not a great fit for tasks in the urgent and important quadrant of the Covey Time Management matrix, as there is a possibility that specific tasks could be delayed. However, it is a good fit for tasks that are important but not urgent quadrant of the matrix, as the law of large numbers ensures that the tasks will be attended to with the relative frequency assigned to it by the user.

If the application proves enjoyable to use, it may provide users with an alternative routine with which to respond to cues that would normally trigger a bad habit. It can also provide users with a bright line limiting the frequency with which they engage in bad habits or undesired behaviors.

9. Conclusion

In this paper, I hope to have outlined the process of developing an artifact which has the potential both in its design and implementation aspects to help people change their behavior in a fun way. I acknowledge that at this point in the process there are limitations both in the research and in the application itself. One limitation is that the kernel theory from persuasive systems and decision support systems have not been synthesized to the extent that is ideal. Another limitation is the lack of one-to-one comparisons of this app to other apps which address similar problems. I intend to address this in future research.

In my studying of the various behavior change techniques, I found a dichotomy with an analogue in philosophy. This dichotomy has to do with the question of why. In philosophy, the teleological perspective focuses on goals, purpose, and meaning, while the mechanistic perspective focuses on direct causes and effects. I find that the behavior change techniques often fall into either teleological or mechanistic categories. For example, if one is trying to lose weight, the setting of a written goal would be an example of a teleological behavior change technique due to its focus on purpose, while an action such as moving snacks to a more hidden and hard to reach place is a mechanistic behavior change technique because it focuses on one of the actual causes of unhealthy eating, viz., a habit or craving triggered by seeing a snack. In future research I plan on fleshing this dichotomy out and conducting experiments to find out whether teleological behavior change techniques, mechanistic behavior change techniques, or a combination of the two is the most effective.

The characterization of bad behaviors as rewards also results in the limitation of the types of bad behaviors the application is able to address. The application is well-suited only to bad behaviors that users want to cut back on, not give up entirely. However, I believe this characterization is beneficial on the whole, as it lends the app it's intermittent reinforcement aspect, making it bear some similarity to gambling machines, which are an example of what Csikszentmihalyi designates as "escaping backward" or escapism which leads to destructive or regressive ends for the individual engaged in it (Csikszentmihalyi, 1993; Csikszentmihalya, 2004; Schull, 2012). By incorporating gaming aspects into a persuasive system informed by decision sciences literature, the app can help users escape forward, and achieve a state of flow while engaged in productive activities on a path of continuous improvement.

10. References

Allen, D. (2015). *Getting things done: The art of stress-free productivity*. Penguin.

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design science research contributions: Finding a balance between artifact and theory. *Journal of the Association for Information Systems, 19*(5), 358-376.

Baumeister, R. F., & Tierney, J. (2012). *Willpower: Rediscovering the greatest human strength.* Penguin.

Biggerstaff, T. J. (1989). Design recovery for maintenance and reuse. *Computer, 22*(7), 36-49.

Carey, R. N., Connell, L. E., Johnston, M., Rothman, A. J., De Bruin, M., Kelly, M. P., & Michie, S. (2019). Behavior change techniques and their mechanisms of action: A synthesis of links described in published intervention literature. *Annals of Behavioral Medicine, 53*(8), 693-707.

Csikszentmihalyi, M. (1993). *The evolving self: A psychology for the third millennium (Vol. 5)*. New York: HarperCollins Publishers.

Csikszentmihalyi, M. (2004). *Good business: Leadership, flow, and the making of meaning.* Penguin.

Covey, S. R. (2020). *The 7 habits of highly effective people.* Simon & Schuster.

Danziger, S., Levav, J., & Avnaim-Pesso, L. (2011). Extraneous factors in judicial decisions. *Proceedings of the National Academy of Sciences, 108*(17), 6889-6892.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, *13*(3), 319-340.

Deveci, M., Mishra, A. R., Gokasar, I., Rani, P., Pamucar, D., & Özcan, E. (2022). A decision support system for assessing and prioritizing sustainable urban transportation in metaverse. *IEEE Transactions on Fuzzy Systems, 31*(2), 475-484.

Duhigg, C. (2012). *The power of habit: Why we do what we do in life and business* (Vol. 34, No. 10). Random House.

Eom, S., & Kim, E. (2006). A survey of decision support system applications (1995–2001). *Journal of the Operational Research Society, 57*, 1264-1278.

Ferrara, G., Kim, J., Lin, S., Hua, J., & Seto, E. (2019). A focused review of smartphone diet-tracking apps: Usability, functionality, coherence with behavior change theory, and comparative validity of nutrient intake and energy estimates. *JMIR mHealth and uHealth, 7*(5), e9232.

Gill, T. G., & Hevner, A. R. (2013). A fitness-utility model for design science research. *ACM Transactions on Management Information Systems (TMIS), 4*(2), 1-24.

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly, 37*(2), 337-355.

Hamari, J., & Eranti, V. (2011, September). Framework for Designing and Evaluating Game Achievements. *Digra conference* (Vol. 10, No. 1.224, p. 9966).

Hamari, J., Koivisto, J., & Sarsa, H. (2014, January). Does gamification work?--a literature review of empirical studies on gamification. *2014 47th Hawaii international conference on system sciences* (pp. 3025-3034). Ieee.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems, 19*(2), 87-92.

Heylighen, F., & Vidal, C. (2008). Getting things done: The science behind stress-free productivity. *Long Range Planning, 41*(6), 585-605.

Hollis, J. F., Gullion, C. M., Stevens, V. J., Brantley, P. J., Appel, L. J., Ard, J. D., ... & Weight Loss Maintenance Trial Research Group. (2008). Weight loss during the intensive intervention phase of the weight-loss maintenance trial. *American Journal of Preventive Medicine, 35*(2), 118-126.

Horsch, C. H., Lancee, J., Griffioen-Both, F., Spruit, S., Fitrianie, S., Neerincx, M. A., ... & Brinkman, W. P. (2017). Mobile phone-delivered cognitive behavioral therapy for insomnia: A randomized waitlist controlled trial. *Journal of Medical Internet Research, 19*(4), e70.

Konrad, A. (2021). Monday.com shares jump At IPO, minting a new cloud software billionaire in Israel. Retrieved from https://www.forbes.com/sites/alexkonrad/2021/06/10/monday-shares-jump-at-ipo-minting-new-cloud-software-billionaire-in-israel/?sh=1e5129125be8

Legaki, N. Z., Xi, N., Hamari, J., Karpouzis, K., & Assimakopoulos, V. (2020). The effect of challenge-based gamification on learning: An experiment in the context of statistics education. *International Journal of Human-Computer Studies, 144*, 102496.

Liu, S., Duffy, A. H., Whitfield, R. I., & Boyle, I. M. (2010). Integration of decision support systems to improve decision support performance. *Knowledge and Information Systems, 22*, 261-286.

Liu, D., Santhanam, R., & Webster, J. (2017). Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS Quarterly, 41*(4), 1011-1034.

Masicampo, E. J., & Baumeister, R. F. (2011a). Unfulfilled goals interfere with tasks that require executive functions. *Journal of Experimental Social Psychology, 47*(2), 300-311.

Masicampo, E. J., & Baumeister, R. F. (2011b). Consider it done! Plan making can eliminate the cognitive effects of unfulfilled goals. *Journal of personality and social psychology, 101*(4), 667-683.

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013, October). Do points, levels and leaderboards harm intrinsic motivation? An empirical analysis of common gamification elements. *Proceedings of the First International conference on gameful design, research, and applications* (pp. 66-73).

Michie, S., Richardson, M., Johnston, M., Abraham, C., Francis, J., Hardeman, W., ... & Wood, C. E. (2013). The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Annals of behavioral medicine, 46*(1), 81-95.

Miltenberger, R. G. (2016). *Behavior modification: Principles and procedures*. Cengage Learning.

Mullarkey, M. T., & Hevner, A. R. (2019). An elaborated action design research process model. *European Journal of Information Systems, 28*(1), 6-20.

Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. *Flow and the Foundations of Positive Psychology* (pp. 239-263). Springer, Dordrecht.

Oinas-Kukkonen, H., & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the association for Information Systems, 24*(1), 485-500.

Prat, N., Comyn-Wattiau, I., & Akoka, J. (2015). A taxonomy of evaluation methods for information systems artifacts. *Journal of Management Information Systems, 32*(3), 229-267.

Rogers, B. (2014). Andrew Filev's Wrike wants to bring project management and collaboration to the masses. Retrieved from https://www.forbes.com/sites/brucerogers/2014/10/15/andrew-filevs-wrike-wants-to-bring-project-management-and-collaboration-to-the-masses/?sh=7f09f1765dd6

Schüll, N. D. (2012). *Addiction by design*. In Addiction by Design. Princeton University Press.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS Quarterly, 35*(1), 37-56.

Sengupta, A., Dutta, K., Beckie, T., & Chellappan, S. (2020). Designing a health coach-augmented mHealth system for the secondary prevention of coronary heart disease among women. *IEEE Transactions on Engineering Management, 69*(6), 3085-3100.

Sutton, R. T., Pincock, D., Baumgart, D. C., Sadowski, D. C., Fedorak, R. N., & Kroeker, K. I. (2020). An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ digital medicine, 3*(1), 17.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Torning, K., & Oinas-Kukkonen, H. (2009, April). Persuasive system design: state of the art and future directions. *Proceedings of the 4th International Conference on Persuasive Technology* (pp. 1-8).

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a framework for evaluation in design science research. *European journal of information systems, 25*, 77-89.

Wang, X. T., & Dvorak, R. D. (2010). Sweet future: fluctuating blood glucose levels affect future discounting. *Psychological Science, 21*(2), 183-188.

Zhai, Z., Martínez, J. F., Beltran, V., & Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture, 170*, 105256.