January 2021

Maximizing User Engagement In Short Marketing Campaigns Within An Online Living Lab: A Reinforcement Learning Perspective

Aniekan Michael Ini-Abasi
Wayne State University

MAXIMIZING USER ENGAGEMENT IN SHORT MARKETING CAMPAIGNS WITHIN AN ONLINE LIVING LAB: A REINFORCEMENT LEARNING PERSPECTIVE

by

ANIEKAN MICHAEL INI-ABASI

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2021

MAJOR: INDUSTRIAL ENGINEERING

Approved By:

_____

Advisor Date

_____

_____

_____

_____

DEDICATION

I dedicate this dissertation to my Lord and Savior Jesus Christ,

and to my family, especially my wife, Margaret, and my daughter,

Miracle, for their unwavering support through the years of my studies.

In addition, I dedicate this work to my global network of prayer eagles for the

never-ending prayers, support, and encouragement.

ACKNOWLEDGEMENTS

First and foremost, I am eternally grateful to the LORD God for His divine guidance and blessing, through which many wonderful people have been brought to help during my research.

I owe special thanks to my academic advisor, Professor Ratna Babu Chinnam, for his unwavering support, encouragement, and insightful advice throughout my doctoral studies. His interest and guidance gave me much-needed confidence to take up this research in the first place. During one of my case study presentations he was quick to see the potentials and possibilities beyond my immediate context, and his guidance helped me in all facets of the research and writing of this dissertation. For over three years we met bi-weekly to discuss a myriad of ideas, out of which came the novel approach that was used in this work. I could not have imagined having a better mentor for my doctoral research.

I would also like to thank Dr. Ming Dong of the Computer Science department, Dr. Yanchao Liu, and Dr. Murat Yildirim for serving on my dissertation committee and for providing insightful advice and constructive criticism for my work. Their invaluable comments and instruction have broadened and deepened my knowledge of my research topic. My special gratitude also goes to Dr. Mohammad Amini for his assistance with the neural network architecture and the state space representation.

Finally, I wish to appreciate my wife, Margaret, and my daughter, Miracle, for their endless love, understanding and support throughout the course of this research. Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Background: User Engagement in Online Systems

User engagement has emerged as the engine driving online business growth. Facebook has pay incentives tied to engagement and growth metrics [119]. Leading corporations are turning to recommender systems as the tool of choice in the business of maximizing engagement. LinkedIn reported a 40% higher email response with the introduction of a new recommender system [81]. At Amazon 35% of sales originate from recommendations, while at YouTube 60% of the clicks on the home screen are on the recommendations [3]. Netflix reports that '75% of what people watch is from some sort of recommendation' [81].

Overall engagement with ecommerce and social media platforms has been uneven at best. As of August 2020, user engagement with Twitter has soared 78% while Facebook has declined by 9% compared to 2018 [165]. With email, the engagement trend has been downward in the ten-year period from 2010 to 2019, even as user growth is projected to reach 4.6 billion by 2025 [134]. However, despite this downward trend, leading brands are turning to email as the medium of choice for reaching and engaging with customers [118]. A recent study shows that 81% of small and medium-sized businesses (SMBs) now rely primarily on email for customer acquisition and retention [54].

The quest for effective methods and tools to maximize user engagement has intensified. With the goal of boosting engagement across all digital touchpoints, researchers and practitioners see personalization as a primary way to achieve this objective [76]. Several approaches have been used to personalize the messages delivered to users at various stages of the customer journey, considering their interests, preferences, behavior, and profiles [111]. Personalized news

recommendation and online advertisements have been a staple for years [69]. On ecommerce platforms, recommender systems deliver personalized pages based on users' search, navigation, and purchase histories. Similarly, on social media such as Facebook and Instagram, advertisements are dynamically served to individual users based, not only on their personal likes, comments, and posts, but also by taking into account the engagement behavior of similar others [148].

While machine learning methods have traditionally been used to build recommender systems with the twin objective of delivering personalized content and boosting user engagement, there is a heavy price to pay in terms of the amount of data that current algorithms expect [41], and the huge "state" (refers to the state of the relationship with the individual user, including past history) and "action" spaces (refers to the set of actions available for engagement) required [49]. Even small-scale recommender systems are characterized by high-dimensional state and action spaces [41], which can present serious computational challenges for traditional algorithms [49, 72, 184]. The choice of available algorithms is limited to those that can learn from large, previously collected datasets, and have so far not been successful at addressing most practical real-world problem settings because of the steep cost of data acquisition in a live setting [17].

1.2 Research Problem

The purpose of this study is to deepen our knowledge on the personalization of persuasive requests to increase their impact on user engagement in a data-efficient real-world setting. The main research question addressed by this thesis is: *"To what extent can a learning agent be used to personalize a persuasive request for maximum user engagement in a data-efficient setting?"* The question is addressed in two parts. The first part of this thesis is focused on questions concerning the effect of "social influence" principles on user engagement [147]:

1. How can a user's susceptibility to social influence principles be learned over a short campaign life cycle?
2. Can a user's response to influence principles be measured and subsequently exploited to increase her level of engagement with future messages?

Based on the knowledge gained about the impact of social influence principles on user engagement behavior, few more questions arise:

1. How do we apply a learning algorithm in the personalization of a persuasive request in a data-efficient setting, given the constraint of short campaigns and product life cycles?
2. How do we reduce the state space given that many practical real-world problems have large and continuous state and action spaces?
3. Should users be sorted and characterized explicitly within the state space or would a universal policy be sufficient for all categories/segments of users?

This research is focused on personalizing message delivery to individual users. Building on the idea of users' predisposition to respond to specific social influence principles of persuasion, we explore how email messages could be customized to individual users to maximize engagement. While the concept of personalizing user experience through item recommendation is already well-established in ecommerce, online advertising and social media, a review of current practices shows that tailoring the *means* of persuasive attempts to user responses has been largely neglected [90, 13, 87]. Unlike the aforementioned domains where vast datasets have been accumulated over long time periods, the settings and scope of email campaigns can vary widely from short campaigns with short product life cycles to campaigns extending over longer periods featuring a variety of products and services. To further complicate matters, many offers might require brand new campaigns in as little as seven days, making offline training much more challenging.

Focusing on small to medium-sized ecommerce enterprises, our problem setting features a small user base, and short-term campaigns promoting products with short sales cycles using email messages. Typical of commercial practice, each campaign starts off as new and typically lasts from seven to thirty days. We leverage a hybrid approach, combining domain knowledge of social influence strategies with data-driven techniques of deep learning to design and present the offers to the user. Personalization of persuasive requests is structured as a "sequential decision-making" problem. Traditionally, the most effective method for solving such problems is through Markov Decision Processes (MDP). Due to the short, uneven duration of some marketing campaigns coupled with observed variability in the engagement behavior of users, the MDP "model" underlying the campaign is unknown and must be learned from available data. Reinforcement Learning (RL), a technique within the broader domain of machine learning, is a proven method for learning effective policies in a sequential decision-making setting. For our purpose, we use Q-learning, one of the most popular reinforcement learning algorithms [178].

In the standard RL framework, an agent learns to properly interact with the environment to maximize the expected total return. Since reinforcement learning can consider long-term reward, it holds the promise of improving users' long-term engagement with any online platform [109]. RL has recorded some successes in a few highly publicized domains, particularly on research problems with high dimensionality and rich environmental data. However, much of the research advances are hard to leverage in real-world systems [79, 50].

In this research we propose a model-free and off-policy Deep RL method, which can train with limited training datasets. Learning convergence is achieved by lowering the discount factor within the algorithm. To reduce the dimensionality of the state space, an aggregation scheme known as the exponential weighted moving average (EMA) is employed. With the underlying

intuition that individual users have a predisposition to respond to different social influence principles, and while their responses may change over time, their observed behavior when presented with a persuasive request could be strongly predictive of future engagement behavior [3, 90].

The first part of this thesis – the insight generation section – is focused on heuristically leveraging the principles of persuasion in specific use cases aimed at increasing aggregate customer engagement value during three real-world marketing campaigns, without the benefit of RL agents. The second part proceeds to design, train and deploy a Deep Q-Network (DQN) agent in a non-stationary environment characterized with limited datasets and short iterations within episodes [50]. The DQN approximates the state-action value function through a deep neural network. The method typically requires many episodes and a large amount of data to converge. However, in this study, agent interaction with the environment is limited. The dataset is not truly representative of the actual state space, as a result the agent suffers from high estimation variance. To reduce this variance and ensure that the Q value does not get too large, a penalty term is added to TD error for training the DQN. A reduced Q value is ideal for short trajectory lengths and improves efficiency where only limited interactions with the environment is possible, in line with recent studies [7]. This novel approach is presented in the methodology section of this thesis in which persuasive interventions are designed, implemented, and applied to identify the persuasion principle(s) most likely to increase user response and engagement with the next campaign message.

1.3 Outline

Figure 1.1 shows a graphical overview of the thesis structure. Chapter 1 gives a background of the problem that is addressed and states the contribution of this thesis. Chapter 2 describes the

current state of user engagement with various online messaging channels, highlighting the psychological foundations, and reviewing the streams of literature relevant to digital behavior. These two chapters together provide an overview of the current state of user engagement and the various approaches employed for incremental gains. Chapter 3 – *insight generation* – reports on several experimental studies which explore the impact on customer engagement value across user groups under repeated exposure to different persuasion principles and suggests the need to personalize influence attempts at the individual user level. Chapter 4 presents a generalized Deep RL algorithm, implements simulation studies to measure the performance of the algorithm, and conducts a real-world case study using a live marketing campaign to demonstrate its ability to significantly increase the level of user engagement, compared with the performance of a human expert and a randomly generated control. In Chapter 5, implications of the current findings, especially on systems designed to maximize user engagement with online messaging channels, are discussed.

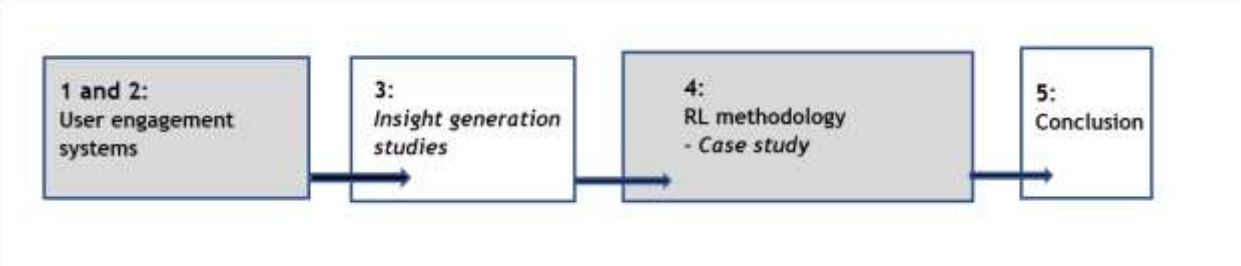

Figure 1: Graphical representation of thesis structure showing the progression from heuristic insight generation studies to RL methodology

1.4 Key Contributions

The overall aim of this thesis is to advance the design of successful persuasive technologies by introducing a data-efficient method to personalize the means by which a persuasive message is presented to the user. The study adds to the current literature by extending our understanding of

human click behavior, enhancing the methods used to understand and capture these behaviors, and advancing technology.

1.4.1 Contributions to our Understanding of Human Click Behavior

There are certain behaviors that are universal for all human beings, one of these being our tendency to respond to psychological principles of persuasion to varying degrees [92, 61, 35, 140]. The next, and possibly more significant behavior, is the seemingly effortless manner in which people can reach important conclusions and make accurate decisions based on limited data [6]. Unlike most data-driven systems that rely on mountains of data to predict likely customer choices and preferences, our study draws from decades of studies on human behavior to guide the design and training of the algorithm in its task of recommending a candidate from a catalog of persuasion principles to drive higher user engagement. The findings of this study could help motivate a shift from an obsessive focus on large-scale, data-hungry systems to making more effective use of "thin slices" of behavioral data in combination with data-driven methods. This hybrid approach is becoming a subject of growing interest, especially for systems designers targeted at optimizing user engagement with small to medium businesses (SMBs), who sometimes lack the critical data infrastructure for current algorithms to work successfully.

1.4.2 Contributions to Methodology

Most online persuasion and personalization attempts to increase user engagement are by-and-large based on the examination of effects over a group of people. While these average effects are often of interest to researchers and practitioners, they are not necessarily a good summary of the effect of influence processes on individual users. This thesis presents a novel use of deep reinforcement learning paradigm in a limited data setting to learn each individual level response to persuasive appeals, with the goal of subsequently using this knowledge to target the right

principle to the right user for maximum engagement. Our methodology shows that design of an efficient state space representation is a critical success factor in reducing complexity, and avoiding the problem of large, dynamic state space, which poses serious issues for traditional RL algorithms. This thesis contributes to the studies that deal with RL on real world systems and, in particular, focus on improving sample efficiency.

1.4.3 Contributions to Technology

As a final contribution, this thesis details the refinements we made to achieve convergence by lowering the discount factor, as a guide to implementers of deep learning algorithm. Specifically, we show by proof of theorem that for the traditional DQN algorithm to handle sparse environment's data efficiently, the discount factor must be reduced. This is in line with current research thinking [7]. Moreover, we identify and present the hyperparameters with the most impact on algorithm performance to aid designers tasked with optimizing user engagement across all digital touchpoints.

CHAPTER 2: RELATED WORK

In attempting to lay a solid foundation for the design of systems to maximize user engagement, researchers often turn to the social sciences that study "persuasion", most notably psychology [92, 61, 35, 140]. A vast array of persuasion principles has been utilized to nudge and influence users to deepen their engagement with online messages, content, ads, apps, products, and platforms, ranging from six principles [34] to forty strategies [61], to sixty-four groups [92], and up to a hundred distinct tactics [140]. The effectiveness of applying persuasion principles to boost engagement has been demonstrated in ecommerce [90], in health and wellness [34, 147], and in social media [152] among others.

In the now-famous 2007 "Facebook Class", psychologist BJ Fogg pushed his students to design and launch apps at break-neck speed using the Fogg Behavior Model. At the end of the 10-week academic term, the applications built by the students had engaged over 16 million users on Facebook. A few weeks after class ended, the total number of engaged users had grown to 24 million. One graduate later redesigned his class app and subsequently cofounded Instagram, the photo-oriented social sharing platform acquired by Facebook in 2012 for a billion dollars [34, 153].

2.1 Persuasion, Engagement and Recommenders

Fogg defined persuasion as a non-coercive attempt to change attitudes or behaviors and proposed "hot triggers" – recommendations that link immediately to desired outcomes [62, 164]. The Fogg behavioral model, which identifies and defines the factors that drive user engagement behavior, traces its roots to classical influence models and theories such as theory of motivation [116], cognitive dissonance [57], attribution theory [73], expectancy theory [175], operant

conditioning [161], social cognitive theory [10], heuristic-systematic model [28], the elaboration likelihood model [132], transtheoretical model / stages of change [133, 110], resistance & persuasion [94], theory of reasoned action / planned behavior [59], and self-determination theory [146].

2.2 Universal Principles of Influence

Prof. Cialdini [35, 34, 36] posits that the power of persuasion techniques is governed by six fundamental psychological principles that direct all human behavior. Correctly applying these principles enables individuals to influence others, induce compliance, and ultimately lead people into agreement and engagement. Multiple implementations of these principles or strategies in both laboratory and field experiments have resulted in more product purchases [64], higher sales [123], and increased product ratings [191]. The six principles include *authority*, *consensus (social proof), commitment and consistency*, *liking*, *reciprocity and scarcity*.

According to the *authority* principle, people are inclined to follow recommendations and suggestions originating from authority figures such as doctors, teachers, coaches, etc. as demonstrated in classical social influence experiments [55, 17, 34]. Individuals tend to feel an obligation to comply with those who are in real or perceived authority positions. *Consensus* (or social proof) holds that individuals view a behavior as correct in a given situation to the degree that they see others performing it. Such individuals who observe multiple others manifesting a particular belief or behavior are more likely to believe and behave similarly [49, 35, 54, 66]. The *consistency and commitment* principle refers to people's desire to act consistently in line with their actions in the past [34] as a way to minimize cognitive dissonance [57], even if they had merely declared their intentions publicly [47]. Once we make a choice or take a stand, we will encounter personal and interpersonal pressures to behave consistently with that commitment [60].

With the *liking* principle, individuals prefer to comply with requests made by those they know and like. When a request is made by someone we like, we are more inclined to act accordingly [34], especially if there is perceived interpersonal similarity [75, 63]. Individuals prefer to comply with requests made by those they know and like.

Where people are inclined to repay a favor, or feel indebted to others in anyway, they tend to comply with persuasive requests to even out the score [68, 35]. According to this rule or norm, one “should try to repay”, in kind, what another person has provided us. This is the principle of *reciprocity* in action. Finally, *scarcity*, whether real or assumed, affects people’s attitudes favorably and increases the chance of purchase [179, 78, 82, 123]. According to this principle, opportunities seem more valuable when they are less available. Scarcity induces compliance in large part because it threatens our freedom of choice ("if I do not act now, I will lose the opportunity to do so") [87]. Many studies draw upon classical research on social influence to explain the profound effect of scarcity on user engagement, such as reactance theory [18], personal equity theory [158], and commodity theory [20], which states that scarce products are more highly desirable because the possession of such products engenders feelings of personal uniqueness or distinctiveness. Some or all these universal principles are currently deployed in the design of persuasive platforms [88, 148].

Another related area of work is in behavioral economics, the study of how psychology – thoughts and emotions – influences and affects economic decision-making. Nobel Prize-winning research defining cognitive biases, heuristics, and prospect theory [84], ergonomics [151], bounded rationality and satisficing [160], nudge theory and decision making [172] have recently become a key component in the design of persuasion and engagement platforms.

User engagement is operationalized through a combination of persuasive technologies and recommender systems. While persuasive technology defines a system designed to change attitudes or behaviors of the users through persuasion and social influence, a recommender system refers to a system which automatically selects personally relevant product or information for users based on their preferences [141, 2]. Whereas persuasive technologies came out of Silicon Valley in the 1990s, recommender systems emerged from information filtering research and nudges from behavioral economics [65, 138, 162].

2.3 Traditional Recommender Systems

Increasingly, the difference between persuasive technologies and recommender systems is blurring, with users more likely to experience and engage with recommender systems as "persuasion platforms" [35]. The systems have been credited with boosting engagement in all dimensions, whether measured as explicit feedback such as ratings, reviews, likes, or implicitly with clicks, product searches or website visits. Conventional recommendation methods can be grouped into three categories. *Content-based methods* [77, 95, 127] are focused on the user profile, and built around historical activity such as likes (on Facebook), keyword searches (on Google), clicks and visits to product pages and websites. The recommender selects items that are more similar to the user's profile and historical preferences. It can capture the specific interests of a user and recommend niche items that very few other users are interested in but has limited ability to expand beyond users' interest.

In contrast, *collaborative filtering (CF) methods* [1, 115, 136] usually make predictions by utilizing the preferences of similar users. Among other disadvantages, the standard matrix factorization approach widely used in CF applications suffer from the "cold start problem", where a new item that has not yet been rated cannot be recommended. *Hybrid methods* [43, 104, 108]

combine the advantages of the two groups of methods using a variety of models such as factorization machines, which merges matrix factorization with regression [136], and neural collaborative filtering, which contains a framework for learning the functional relationship modeled by matrix factorization with a neural network [70]. With the recent extension and integration of previous models, new deep learning models such as DeepFM, Wide and Deep, Facebook's DLRM have shown much superior performance than all three previous categories due to the capability to simultaneously learn high and low-order feature interactions [70, 29, 125].

2.4 Reinforcement Learning Based Recommender Systems

There is a growing body of work focused on applying reinforcement learning (RL) to the recommender problem. As a result, RL has recently achieved impressive milestones in such areas as news recommendation [190], 2018), online advertising [26, 188], and YouTube recommendation [31]. RL seeks to build software agents that interact with an environment to maximize some notion of long-term reward. One stream of research formalizes the recommendation problem as a "multi-armed bandit" (MAB), in its simplest form, a model for the exploration/exploitation trade-off [173], where the agent needs to balance trying out actions to learn their associated rewards and selecting only those actions that lead to high rewards [169, 177, 185, 102]. Beyond the simplistic bandit, the contextual multi-armed bandit (cMAB) has access to additional information, the context, which might influence the reward associated with certain actions [103, 171, 105]. There is further research combining bandits with matrix factorization [27] for the purpose of modeling more complex reward functions and user/item relationships.

Beyond contextual bandits, there is even more substantial research interest in Markov Decision Processes (MDP), a classical formalization of sequential decision making which is a mathematically idealized form of the reinforcement learning problem. In contrast to the bandit-

based methods, MDP-based methods not only capture the reward of current iteration, but also the potential reward in the future iteration [124, 112, 189].

Researchers approach RL methods from two perspectives: *model-based* or *model-free*. Algorithms which explicitly learn system models and use them to solve MDP problems, are model-based methods. They include popular algorithms such as the Dyna [167], Prioritized Sweeping [124], Q-iteration [25], Policy Gradient (PG) [180], and the variation of PG [13, 85]. The model-free methods ignore the model and instead focus on figuring out the value functions directly from the interaction with the environment. Some examples include Q-learning [98], SARSA [145], LSPI [100], and Actor-Critic [96].

RL algorithms which first find the optimal value functions and then extract optimal policies are "value iteration" methods, such as Dyna, Q-learning, SARSA, and DQN [120]. The alternative approaches are policy search methods which solve MDP problems by directly searching in the space of policies, such as "policy iteration" (PI) algorithms [113, 13, 115, 45]. There are a series of algorithms, which use the PI to search in the policy space, and at the same time estimate a value function. The important class of these methods are Actor-Critic (AC) and its variations [96, 126, 45, 130, 131, 15].

Deep neural networks provide rich representations that enable many RL algorithms to perform effectively. However, it was previously thought that the combination of RL algorithms with deep neural networks was fundamentally unstable. Therefore, a variety of solutions have been proposed to stabilize the algorithm [135, 120, 122, 71, 154]. These approaches are based on the idea that the sequence of observed data encountered by the agent, though non-stationary, is strongly correlated with the updates. By storing the agent's data in an experience replay memory, the data can be batched [135, 154] or randomly sampled [120, 122, 71] from different time-steps.

Aggregating over memory in this way reduces non-stationarity and decorrelates updates, but at the same time limits the methods to "off-policy" reinforcement learning algorithms.

Deep RL algorithms have achieved unprecedented success in challenging settings, but with several drawbacks, among them high-dimensional state and action spaces. Recommender systems deal with large state and action spaces, and this is particularly exacerbated in industrial settings. The set of items available to recommend is non-stationary and new items are brought into the system constantly, resulting in an ever-growing action space with new items having even sparser feedback. Further, user preferences over these items are shifting all the time, resulting in continuously evolving user states. There are several recent works focused on addressing these challenges [41]. Dulac-Arnold et al. present a solution based on generating a vector for a candidate action and then performing nearest neighbor search to find the closest real action available [17]. Zahavy et al. uses a contextual bandit to eliminate irrelevant actions [16]. Osband et al. learns an ensemble of Q-networks and applies Thompson Sampling to drive exploration and improve sample efficiency [126]. Finn et al. addresses sample efficiency by using "few-shot" learning to learn about tasks within a distribution, quickly adapting to solve a new in-distribution task not previously seen [58]. Another approach to improving sample efficiency is to learn ensembles of transition models and use various sampling strategies from those models to drive exploration [74, 33, 24].

Although RL and Deep RL have been successfully applied to many recommender problems, the algorithms only converge when large training datasets are available. There is growing research focus on sample-efficient (a.k.a. data-efficient) exploration and the use of augmented data. Unlike much of the current research, our study requires an algorithm to be both sample-efficient and performant in its operation. While an MDP framework with continuous state is proposed, an exponential moving average (EMA) scheme is used to capture the information on

state and action over time and to reduce dimensionality of the state space. In contrast to most recommender systems which only suggest items to users, this study focuses on how to frame the message to each user with the right persuasion principle to elicit maximum engagement, irrespective of the end goal.

Current recommender systems have access to large volumes of historical data amassed over long time periods, unlike our setting which features short campaign cycles of a few weeks' duration, traffic fresh from a social media ad campaign, with no prior historical data on user preferences. Therefore, the policy is hard to learn. Different from the literature, we propose a novel model-free and off-policy Deep RL, which can be trained and deployed in a data-efficient manner. Given this constraint, we turn to the idea of "thin slices" of data, as proposed by Ambady and Rosenthal [6]. In their landmark study, they showed that much inference is possible just by observing "thin slices" of nonverbal behavior (in this case user engagement behavior). Curhan and Pentland [42] applied the idea in a simulated employment negotiation scenario. They found that certain engagement features within the first five minutes of negotiation were predictive of the overall negotiation outcome in the end. Other researchers [129] have demonstrated that thin slices of data can be used to observe a short window of behavior to achieve prediction comparable to observing the entire data stream. It is proposed the same idea applied to user click data in ecommerce engagement settings could be predictive of future engagement behavior.

CHAPTER 3: HEURISTIC APPLICATION OF PERSUASION PRINCIPLES TO INCREASE USER ENGAGEMENT: INSIGHT GENERATION FROM LIVING LAB TRIALS

Online communities provide an ideal platform for researchers to investigate user engagement as well as social mechanisms related to behavior and decision-making in a rich variety of contexts [44]. At the intersection of data and the social sciences lies the potential to enrich our knowledge of individuals, groups, and societies with an unprecedented breadth, depth, and scale [101, 51, 64]. “Living Labs” enable the ability to capture and analyze many facets and dimension of human behavior, communication, and social interaction among members of a target community. They facilitate interventions in the community, with the opportunity to precisely measure their effect – both through implicit means (automatic sensors, clicks) and explicit assessment (e.g., questionnaire, surveys) of individuals and collectives [18, 86, 147, 88, 89].

Although some theoretical discussions differ on the actual definition of a Living Lab, most authors agree that it is a way to involve end-users in innovative research, over a longer period of time, using a combination of different research methods [156]. Living Labs use the participation of end-users to gain better insights into the possibilities and restrictions of innovations [155]. They have steadily grown in value because of the digital revolution and the increased tendency of researchers to extend the research processes beyond the limitations of the closed laboratories toward the highly dynamic environment of "real life" [39]. Long regarded as environments that enable experimentation with real users in their natural contexts, the concept originated from MIT's Media Lab, where it was initially used to observe the living patterns of users in smart homes [144]. There is now an emerging trend to expand the concept to include online research activities, and to enhance innovation, usability of information technology and its applications in the society [14].

Established on four core elements: (1) being a research and development process of

innovation, (2) being a collaboration between multiple stakeholders, (3) taking place in a real-life setting, (4) involving users as co-creators, Living Labs have been applied to develop and test different concepts and innovations in co-creation with users [46]. The methodology delivers a new way of structuring research through validation and testing in real-life contexts. While some emphasize the social innovation aspects of Living Labs [156], other researchers argue that Living Labs, whether physical or virtual, whether social or technical, ultimately help to collaborate for creation, prototyping, validating, and testing of new technologies, services, products, and systems that better meet user needs. The methodology offers a socio-technical infrastructure to support user-centric innovation processes, deliver collaborative platforms where researchers, practitioners and users work together to generate solutions that are rooted in the settings of daily life practices [144]. Given that social innovations often feature open and ill-structured problems, the Living Labs methodology offers a promising alternative to linear and closed modes of problem solving, thus facilitating user-centric and user-driven practices in real-life contexts [14].

An important element in Living Lab research is the study of user engagement with a technology or prototype in their natural environment early in the innovation process and at later stages. Field trials, which can be defined as “tests of technical and other aspects of a new concept, product or service in a limited, but real-life environment” are a method to stimulate the interaction between the technology and users in their real-life environment [8]. Whereas early Living Lab field trials took place in special laboratories, and second-generation Living Labs conducted field trials in a real-world use context, in relatively uncontrolled settings outside of the laboratory, latest generation Living Labs increasingly incorporate major online components [150, 11, 106].

Field trials are one of several approaches to discover and understand how technologies are being used in the wild, with all the socio-technological parameters present [93]. The goal is to explore

the users' understanding, practices, and eventual engagement with the system. Different data collection methods usually implemented include data and sensor logging, interviews, observations, or user reports [22]. The degree of user engagement can provide insights into the effect of social influence on choice, motivation, and behavior of individuals as well as groups.

In this chapter we use three field trials to explore the relationship between user engagement signals and the persuasion principles applied to boost or strengthen them within an online community. In this context, user engagement, primarily measured through clicks, has been construed as providing word of mouth, comments, blog posts, customer ratings and reviews for a product or brand [174]. To facilitate continuous improvement of the offerings presented, users contribute resources such as content, knowledge, skills, and time, recommendations, referrals and many other behaviors influencing the firm and its brands [48]. Users can express their opinion, offer suggestions, post comments, co-create value, collaborate in the innovation process, and become endogenous to the firm [16]. Researchers suggest that user engagement encompass behaviors through which users can make voluntary contributions that have a brand or firm focus but go beyond what is fundamental to the transaction [80].

The literature on the usage of persuasion-based strategies to increase user engagement has been growing. Several works have explored the design space and the application of machine learning to personalize and tailor messages as a way of boosting engagement in such diverse areas as marketing, healthcare, and ecommerce. To address the problems of declining engagement and low adherence, Paredes et al. [159] applied reinforcement learning to recommend stress coping interventions tailored to individual users. Online personalized news recommendation is another active research area where personalization has been used to increase engagement [77, 95, 81, 1, 29]. Study shows that if personalized persuasion principles were embedded in email messages,

overall engagement with such messages would be enhanced [90, 86, 147, 88], which is the primary focus of this paper.

Our approach is aimed at boosting *total* user engagement value within a cohort of online customers through a strategy that allows a human agent to intuitively apply persuasion principles at periodic intervals. As digital markets continue to grow, many businesses are challenged by declining customer engagement, with quantifiable negative impact on revenues and profitability. However, 85% of marketers still see email as a successful channel for achieving the business objectives of customer acquisition, conversion, and retention [139]. To explore user engagement our framework relies on embedding persuasion principles in email messages to cohorts of users in a Living Lab setting. This insights generation chapter studies the *aggregate* increase in responses to social influence strategies. The chapter uses a selection of the social influence strategies, as suggested by a human agent based on heuristics, to investigate the effects of increased user engagement on the revenues of an online bookstore. This set of experiments is an attempt to use persuasion principles to boost engagement beyond "open rates" and "click-through rates" to deeper levels of engagement, as evidenced by users taking further, more definitive action e.g., posting a comment, or purchasing a product [99]. Leveraging three studies, this chapter examines the size of the aggregate impact of persuasion strategies, and the stability of these signals over time. All three studies present selections made by a human agent using a heuristic technique to produce solutions that though not optimal are nevertheless sufficient given the setting. Based on the responses, a broad measure of the effectiveness of each strategy can be obtained, thus setting the stage for a RL agent to make recommendations at the individual user level in the next chapter.

Previous works have proposed persuasion-based systems as a means of optimizing the persuasive appeal and user engagement metrics among various cohorts from students to online

shoppers [90, 86, 147, 88, 89]. However, these studies have been conducted in a one-off setting, without the depth of engagement that can occur with field trials conducted in a Living Lab setting. Ståhlbröst et al. distinguish four aspects that influence user engagement in Living Lab processes and more specifically in online communities. The first aspect is the *process* itself, which includes the timing and the implementation of the research activity. An unfamiliar process can give users the feeling that the study is not sufficiently related to their daily experience. On the other hand, a run-of-the-mill implementation can demotivate test users, and decreasing their willingness to engage and to provide feedback. A second aspect is the (Living Lab) *community*, which includes the participants, the presence of a facilitator, the possibility to getting rewarded, if only emotionally, and the extent to which users are motivated to participate in the community. A third aspect is the quality, timeliness, and appeal of the *content*. The fourth and final aspect is the *platform*, which needs to support, store, monitor and motivate its operations [166].

When looking at the practical organization of field trials, several barriers must be considered. First, participants can adapt their behavior according to what they consider to be important information. Second, social relations can influence the results of a field trial, for example, innovative participants can stimulate other participants to engage. Third, the design of the trial and the way in which the trial is being presented to participants can influence the level of participation. With these barriers in mind, Brown et al. argue that researchers should act as participants instead of controlling the field trial [22].

From a tactical standpoint, user engagement must take into account both explicit and implicit perspectives such as purchases (and post-purchase behavior in the form of posting reviews, feedback and ideas for improvements), knowledge sharing (e.g. sharing experience with others), and co-developing behaviors (e.g., ideas for new products)[80, 99]. Based on a satisficing

heuristic [84, 172], we conduct three field trials to evaluate the engagement signals within the Living Lab. The proposed scheme is practical and can be operationalized in small to medium businesses (SMB) without significant changes to their current messaging practices.

3.1 Contributions

This work allows us to make the following contributions:

1. We present a Social Innovation Living Lab (SILL) for capturing and analyzing the engagement behavior within a SMB-affiliated online community. This structure necessarily incorporates persuasive appeals that boost many dimensions of user engagement beyond just the clicks such as purchases, reviews, and knowledge sharing. Without assuming that all users will respond similarly to persuasive appeals, our approach relies upon the core motives model of social influence for strategy implementation. Our heuristics selection technique successfully increases engagement in two separate field trials. Though these results may not be optimal given the relative lack of rigorous analysis, nevertheless, the results may be deemed sufficient for SMBs looking for a practical method to increase user engagement.

2. We characterize the benefits of employing a persuasion-based messaging scheme under various combinations of long/short campaign duration, number of messages sent, and number of persuasion principles per message. We demonstrate that a longer campaign period may not necessarily lead to higher user engagement but could very well result in overall lower engagement. Experimental trials featuring more than one persuasion principle per message appear to be more effective, yielding up to 37% higher informational and transactional value, accompanied by a 50% reduction in campaign time. We also attempted to quantify the impact of each additional persuasion principle on the observed increase of user response and engagement.

Key insights resulting from our study are the following: 1) Online businesses challenged by declining user engagement could potentially look to persuasion-based messaging to reverse the trend, 2) Careful selection and application of persuasion principles in business communication at the appropriate stages of the customer journey can help cultivate a positive association, reduce uncertainty, and motivate action that results in higher user engagement, 3) Conducting persuasion-based messaging and intervention in an online community setting can help determine which principle could potentially lead to higher engagement, 4) The heuristic approach to persuasion-based messaging is inherently low-risk, requires little to no additional infrastructural investments, and can be productionized within days without disruption to existing systems, 5) Benefits can be realized even in the absence of highly accurate or optimized processes in place, 6) Proposed persuasion-based methodology can potentially sustain, as well as increase, user activity, interest and participation, and 7) This study lays the groundwork and gives added impetus to the need to optimize the strategy selection process by designing a message selection recommender system that takes into account each user's predisposition to respond to specific persuasion principles, in contrast to the mass application of the same principle to all users in the cohort.

3.2 Field Testing within Living Labs

3.2.1 Methodology

This research is based on the high-level framework proposed by Coorevits et al. [40]. Composed of four quadrants along two axes: degree of realism (high vs. low) and phase in the Living Lab project (early vs. late), four "archetypes" of field tests emerge as shown in Figure 3.1 below.

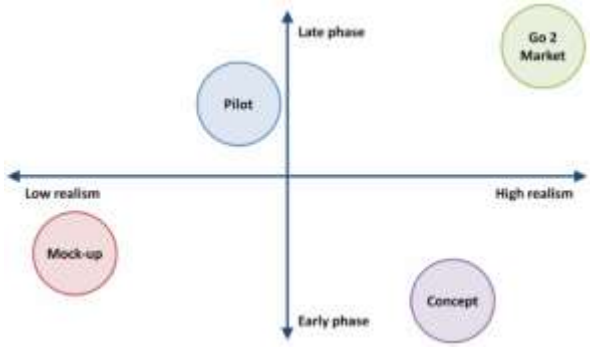

Figure 2: The four types of field tests in living labs *(Source: Coorevits et al, 2018)*

Several studies to validate and fine-tune the framework have been performed using qualitative multiple illustrative case study. Defined as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context [157], case study research has been used to analyze more than 100 living lab projects to identify these four archetypical field tests, and to establish guidelines for best practices [12]. Out of these cases, four field tests have been identified that best matched the four archetypes including: 1) *concept field test*, which help identify the user's problem in the early stages of new product development, 2) *mock-up tests,* designed to gather information about the nature of the interaction and test it before the functional model is built, 3) *pilot field test,* focused on testing the entire system with a subset of users in real-life conditions and can be perceived as the dry-run test of the innovation, and 4) *Go2market field test,* which is mostly used to validate the innovation concept when the maturity is at a higher level. In this quadrant, the focus is on questions relating to product-market fit, willingness-to-pay, engagement, retention, growth, go-to-market strategies, and scaling. Go2market field tests are characterized by a high degree of realism, interactivity, and a large sample size.

The living lab field tests in this study fall under the fourth archetype, and exhibit the following characteristics:

1. *Large-scale and open:* It includes a larger group of test users and everyone who qualifies can participate. This larger group of test users is needed to get a statistical validation of the proposed innovation, potential future roadmap based on adoption potential per target group, and to operationalize the willingness to pay.
2. *Limited to no guidance:* The main focus is to ensure the process can withstand the highly dynamic contextual requirements which sometimes act as a driver or barrier to engagement and, therefore, the test should be as natural as possible, featuring limited involvement of the researchers, with limited-to-no guidelines given to the users regarding how to engage. This also implies the test is less intrusive for the user.
3. *Quantitative:* As the focus is on validation and larger user groups are involved, the methods used will be more quantitative in nature. Questions about "what" and "how many" will be answered during these field tests. Log data from the system and measurements (implicitly and non-intrusively) will take place at several time intervals or when certain events take place to learn about how users behave, their attitudes, and their suggestions for future improvements.

3.2.2 Living Lab Setting & Observations

This section briefly introduces the operations of a Living Lab associated with an online bookstore by analyzing data gathered during three marketing campaigns spanning multiple years (January 12, 2017 - March 31, 2018), involving 9,205 anonymized subscribers generating 122,491 engagement signals. Table 1 summarizes the user engagement data over the 3 campaigns. During each campaign, which lasted between 30 and 90 days, users received one message per day. The messages were short (250 to 400 words), embedded with 1 to 2 of Cialdini's universal principles of persuasion, and delivered via permission-based email to each user's inbox daily at 12:00 A.M.

EST. Deeper, non-traditional metrics of engagement such as replies, ratings, comments, and purchases were preferred over traditional email metrics of opens and clicks.

Starting January 2017, we initiated a Living Lab study conducted with users who had opted in to access further content after placing an order. All members of the community are customers of Beyond Books Publishing, an online retailer of bundled books and programs focused on wellness and behavior change. It offers a curated selection of bestselling books with a unique selling proposition: ordering a featured book during the campaign period comes with an entry pass to a membership site where additional bonuses can be accessed. Such bonuses often include podcasts featuring the book authors, lively discussion forums, and extra chapters that did not make it into publication.

Membership confers status on the users similar to having a backstage pass into a concert. It effectively functions both as an engagement device and an opportunity to offer other merchandise at discounted prices once inside the forum. Each user voluntarily filled out an opt-in form after purchasing a featured item from the Beyond Books website or directly from Amazon (Figure 3.2a, b). Out of nearly 10,000 subscribers, approximately a quarter opt-in to participate in each marketing campaign. The "residence" has a vibrant community life and many virtual ties of friendship between its members who do not share personally identifiable information and have never met in real life. However, they are bonded by a shared passion – holistic wellness and motivational literature. We shall refer to this online residence as the "*Virtual Friends and Family*" community.

Messages were delivered to users through email. For the baseline campaign, the messages were composed and sent without persuasion principles. Over a 90-day period the messages were sent daily for the purpose of motivating and persuading the subscribers to order the featured

product and opt-in to access the membership area. Given that users could engage at any time during the course of the campaign, order the item and obtain the entry pass, the context was completely natural.

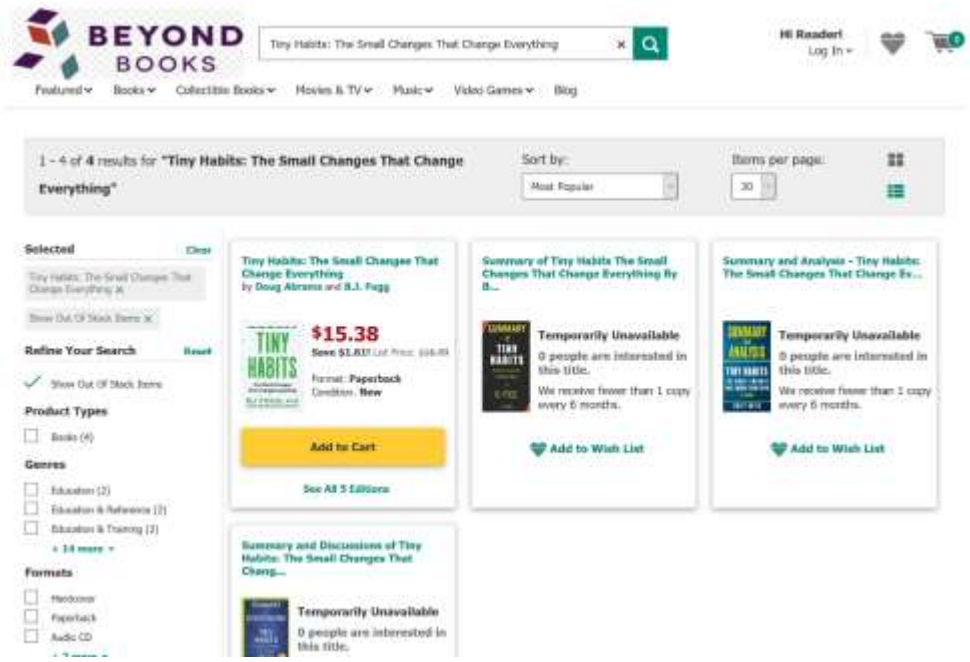

Figure 3: Beyond Books website with featured book

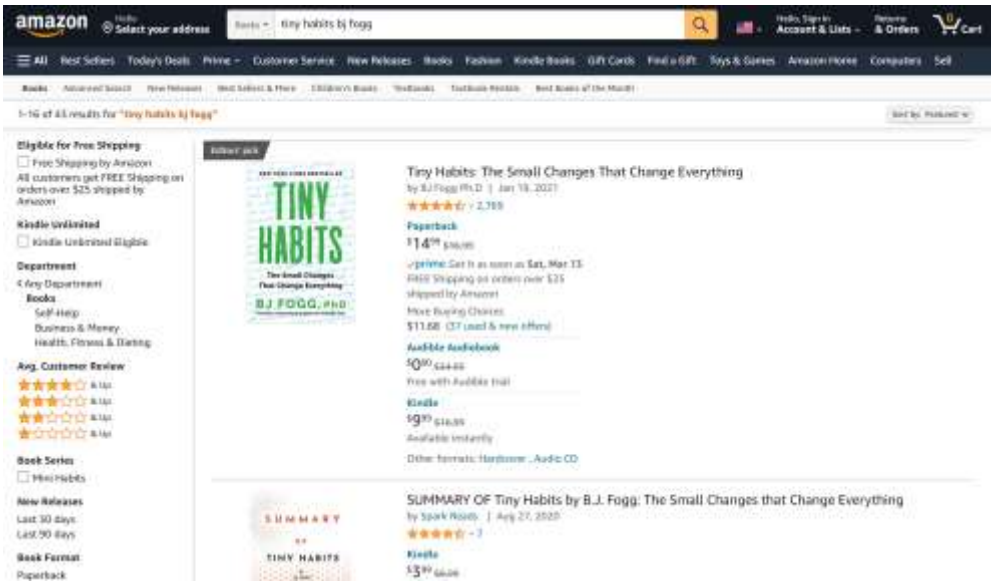

Figure 4: Amazon Storefront with featured book

The messages were composed and delivered using one of three commercially available email service providers – Aweber, Kajabi and Mailchimp. The goal was to start off with a long, customer focused engagement campaign to learn, adapt, and establish a baseline for future field trials scheduled to be conducted in a much more compressed timeframe. A skeletal version of an agile email marketing model was used to build out the entire campaign of 100 messages within a two-week period (Figure 5).

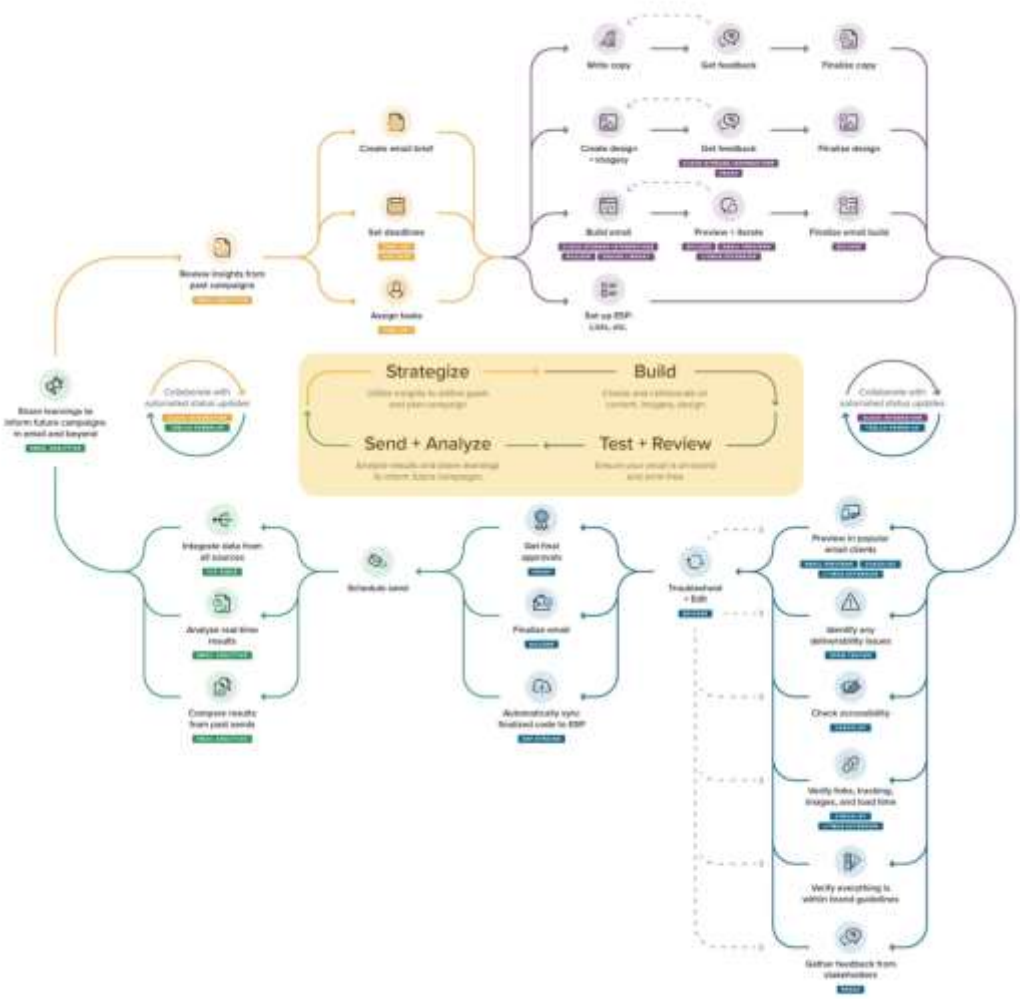

Figure 5: Agile email marketing model (Source: Litmus)

The baseline campaign ran over a much longer period to generate sufficient data to identify and strengthen the weakest links in the Ability Chain, which is made up of five links: time, money, physical effort, mental effort, and routine (Figure 3.4). A new study suggests that the most engaged users would be willing to spend time, money, mental and physical effort, and participate in routine activities [68]. Users most willing to perform a complicated series of steps in order to get or achieve something invariably turn out to be the most active and engaged.

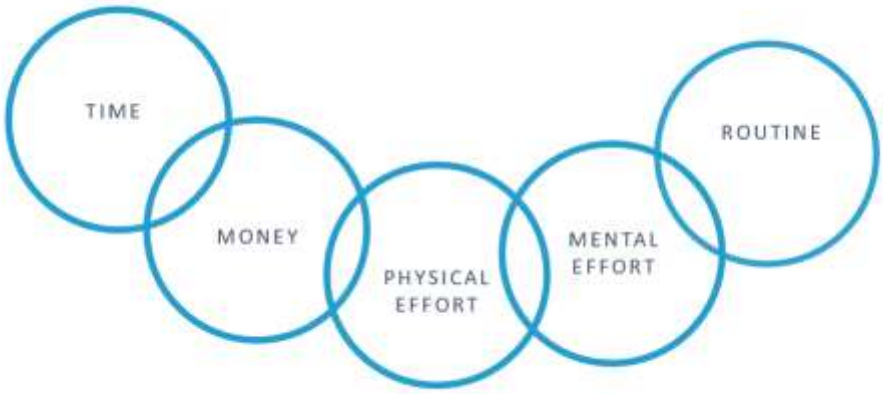

Figure 6: Ability chain of user engagement *(Source: BJ Fogg, 2020)*

During this Living Lab, three different field trials were conducted. Field trial #1 was the baseline case launched in January 2017. Email messages were sent to a list of 9,025 subscribers who had previously opted in to receive short updates and summaries on the latest catalog of motivational and wellness publications, along with reader commentaries from around the world. Each message included a link to an offer to: 1) order the book from Amazon or the company's website, 2) sign up to a members-only forum for extra bonuses, podcasts and free access to experts, fans and hobbyists focused on implementing the ideas expressed in the books. Though users were pre-qualified to join the forum after their order was processed, they were also required to explicitly fill out a double opt-in form. The double opt-in added an additional step to the email subscription

process, requiring the user to verify their email address and confirm interest. By using a double opt-in confirmation method, the chance of spam addresses in the subscriber list was greatly reduced. Email messages were broadcast daily to motivate engagement with the content and with other users in the Living Lab. An icon, or avatar, was used to represent the users in all online discussions without personally identifiable information being disclosed. Through the daily email messages users were encouraged to participate in discussion threads, question-and-answer sessions, webinars, reviews, and to take advantage of other discounted products offered exclusively to members. During the first field trial, which ran for a period of 90 days, 100 email messages were sent, 1,524 users opted in (a 16% response), and 28,102 engagement signals were recorded. The messages were sent with plain subject lines, and without any persuasion principles embedded. A baseline was established against which subsequent trials could be measured.

The Living Lab field trial #2 introduced the persuasion principle of consensus (social proof) into the subject lines of all messages. Campaign duration was reduced from 90 days to 40 days. The message count was cut in half from 100 to 50 emails. 2,730 users opted in (representing a 30% response) and generated 44,189 engagement activities. Living Lab field trial #3 focused on further reducing both the message count and the campaign duration, while adding a secondary persuasion principle (scarcity). This trial ran for the shortest period of just 30 days and featured the least message count (30). 50,200 engagement signals were generated by 2,950 opt-in users, representing a 32% response).

Field trial #3 examined the effects of using multiple influence principles to support a single appeal as opposed to the selection of one specific strategy. The problem of the simultaneous presentation of multiple influence strategies to support a single appeal has been understudied but is valuable to designers of persuasive systems. Only within the marketing literature have serious

attempts been made to address this question but the results are not conclusive. Barry and Shapiro find that using multiple social influence principles can produce adverse effects and hurt compliance [12]. Thus, when a single influence principle is used, it seems more effective than using multiple principles. Falbe and Yukl however arrive at a different conclusion from observing multiple human-to-human persuasion attempts within a company setting: they show that managers who are flexible and use multiple persuasion tactics on the same target to support a single appeal are generally more successful than those sticking to a single tactic [56]. These existing studies are however correlational and thus they do not provide causal evidence. Field trial #3 sets out to experimentally test the effects of multiple persuasion strategies versus a single strategy (field trial #2). Table 1 summarizes the results.

	Baseline 1	Field trial 2	Field trial 3
Campaign period (days)	90	40	30
Message count (daily)	100	50	30
CEV (money)	1,524	2,730	2,950
CIV (effort)	28,102	44,189	50,200
Conversion rate	16%	30%	32%

Table 1: Summary of user engagement data during field trials

This study involved a relatively different subject population when compared to previous studies such as colleagues and co-workers [8], undergraduates [10] and young couples [11]. The *Virtual Friends and Family* community includes a much more diverse and heterogenous subject pool and provides a unique perspective into a user group that has not been traditionally studied in a Living Lab setting – online customers united by a shared passion.

Within the Living Lab approach, some of the challenges faced by researchers include finding motivated long-term users [83] and generating a critical mass of activity to keep them

engaged throughout the trial [97]. One of the reasons that participants were recruited from an existing customer pool was to implicitly activate the influence principle of *consistency and commitment*, which suggests that people's desire to act consistently in line with their actions in the past would most likely attract only committed and passionate users [28, 30, 62, 66].

3.3 Discussion and Conclusion

This study was performed using a comparative case study analysis of three Living Lab field trials. Throughout the study engagement, activities of the users were measured implicitly and anonymously, without the need for questionnaires, surveys, or data tracking tools such as sensors and mobile phones. The main critiques of this method are a lack of scientific rigor, the unstable basis for (scientific) generalization and the difficulty of analyzing huge amounts of data coming from different sources. This study was aimed at generating insights on the possible effect of persuasion principles on engagement within a cohort of users involved in online field trials, instead of aspiring to rigorous statistical generalization. The study touches on many aspects of user engagement from clicks to purchases to community behavior in response to selected principles of persuasion.

To reverse the trend of declining user engagement being experienced by SMBs, with its negative impact on sales, profit, and growth, we propose the incorporation of persuasion principles in their email marketing messages to both prospects and customers alike. The proposed implementation does not require any major modifications to their marketing and transactional messages. Our results suggest that embedding certain principles of persuasion in campaign emails can significantly increase user engagement for an online business (and have a positive impact on revenues) without putting pressure on marketing or advertising budgets. During the study, the store had a customer retention rate of 76% and sales grew by a half-million dollars from the three field

trials combined. The results led the store to implement the persuasion-based messaging strategy in its natural health vitamins and supplements store. The company is also planning to experiment with a more personalized strategy designed to target the users at the individual level with different persuasion principles. The personalized messaging approach can be studied for further operationalization.

First, given the dynamic nature of the online environment characterized by ever-diminishing attention spans, users still appear to respond to influence principles of persuasion when judiciously applied to email campaign messages. While one persuasion principle embedded in the subject line can lead to a significant increase in user engagement, a second principle in the message body can reinforce and boost it even further.

Second, persuasion-based user engagement can enable desirable customer behavior by driving customer loyalty. For instance, when users are given an opportunity to engage with similar others, they tend to be more willing to purchase more products and participate in future events, thus boosting the customer retention rate. In the Living Lab, we see that social influence principles can have observed effects on personal choice and behavior. Once people are embedded in a social fabric (in this case virtual), individual decision-making ceases to be performed in a vacuum. This provides a strong case for social proof in action, which should be harnessed to design social mechanisms that would support positive and desired behavior change. In this setting, the principle of social proof can be utilized for proactively "nudging" the users in the direction of compliance to a persuasive appeal. While different settings and use cases exist, this study provides core ideas coupled with a satisficing approach that could be readily adapted for implementation even by resource challenged SMBs in the online marketplace.

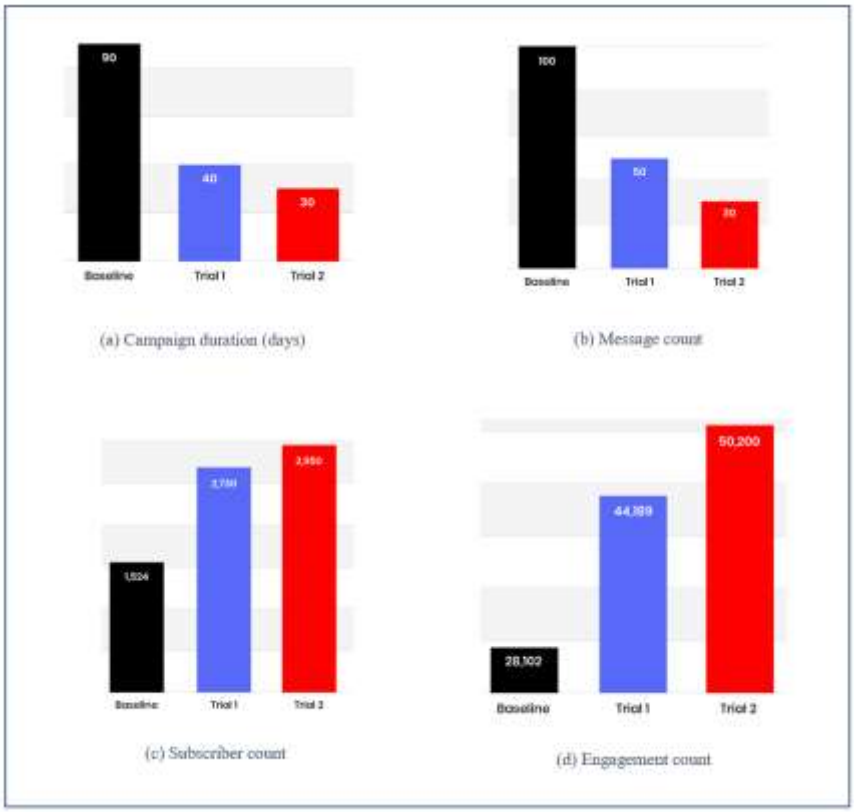

Figure 7: Comparison of Living Lab Field Trial Results

Finally, even though this study is focused on using persuasion-based campaign emails to increase user engagement in an online store setting, it adds to the general body of literature investigating customer acquisition and retention channels that could benefit from higher user engagement (e.g., social, voice, and text messaging), adaptable to industries as diverse as retail, healthcare, and online advertising. Using a heuristic approach to select the principles for the next message, however, has drawbacks which could lead to sub-optimal results.

Nevertheless, insights generated from this study could form the basis for the design and implementation of a persuasion-based recommender to increase *individual* user engagement with email messages. Incorporating persuasion-based thinking and strategies into the implementation

of recommender systems should become a promising area of scientific research and exploration for both industry and academia.

CHAPTER 4: MAXIMIZING USER ENGAGEMENT THROUGH GENERALIZED REINFORCEMENT LEARNING

4.1 Introduction

While small to medium businesses (SMB) struggle with declining clickthrough rates (CTR) leading to overall lower user engagement, large technology brands have witnessed increasingly higher user engagement through the deployment of complex recommender systems. These systems have been credited with driving higher user engagement, sometimes considered as a proxy for customer retention, as well as other measurable business value such as higher sales and revenues, higher click-through rates and dwell times [67, 81, 163]. For many online platforms, where high user activity has been linked to positive business results [4, 32, 91], an increase in user engagement is strongly correlated with the deployment of recommender systems.

With the introduction of recommender systems, studies have variously shown a 6% improvement in revenues at eBay, a 50% higher activity level at a music recommendation site [1], and a 35% lift in sales at an online retailer [102]. With its recommender system, Yahoo! Answers reported a 17% increase in the number of answers and a 10% increase of daily session length [179]. LinkedIn reported a 40% higher email response with the introduction of a new recommender system [180]. A set of studies found that at Netflix, the company's personalization and recommendation service have helped to decrease customer churn by several percentage points over the years, with estimated business value of $1 billion per year [4]. Each of these successful cases invariably involves deploying complex algorithms, running A/B split tests with large sample sizes of millions of users for long periods of time, e.g., several months. Within the context of these large-scale businesses, even small changes in revenue, e.g., 0.5%, can have a significant impact on business results [183].

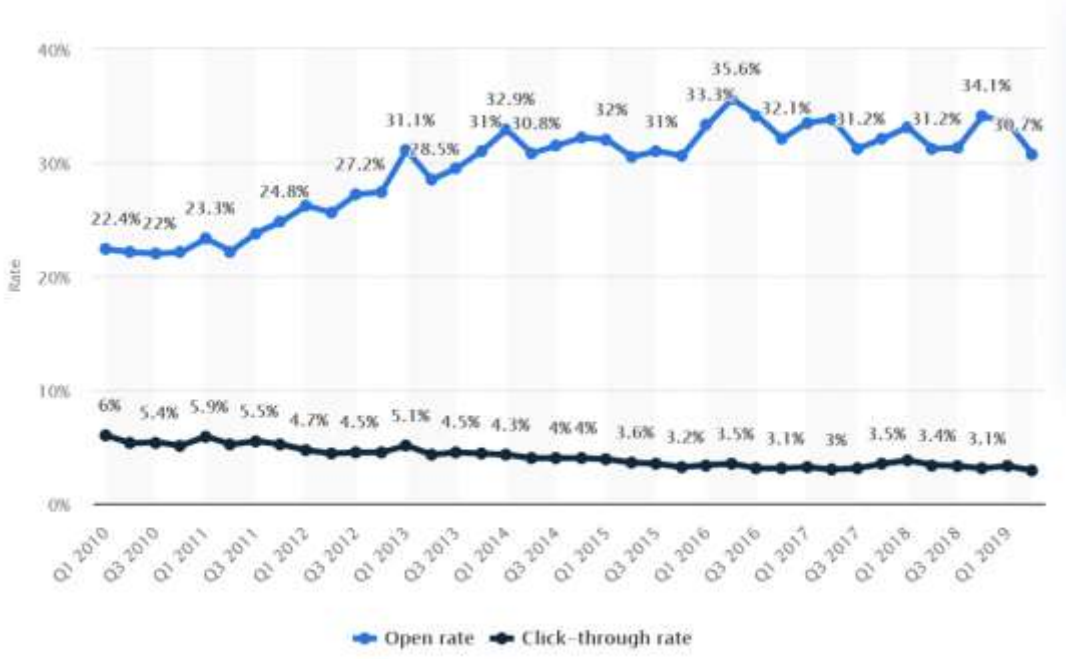

Figure 8: Decade-long decline in click-through rates *(Source: Statista)*

The motivation of this study is the search for a sample-efficient recommender for delivering personalized messages, with the potential of yielding higher user engagement benefits but without the severe implementation challenges of massive recommender systems. Our methodology will be especially useful in the context of SMBs with declining CTR as in Figure 4.1, but without the resources to deploy large, complex systems to help reverse the trend. Unlike leading technology brands, SMBs operate under enormous constraints such as insufficient data, short marketing campaign cycles, and limited technical and financial resources. Studies have shown that recommender systems, in general, are hard to leverage unless a set of practical challenges are addressed. These include the challenge of measuring the business value, algorithm choice, the pitfalls of field tests, and training offline from the fixed logs of an external behavior policy [17, 67].

While there has been significant progress in using recommenders to increase user engagement both in theory and practice, a great deal of the attention has been focused on recommending items to users. For most personalization and recommendation services, the end goal is to present items that the user is most likely to engage with, whether that be a product (Amazon), music (Spotify), movies (Netflix), videos (YouTube), or news (Google). Little attention is paid to personalizing the *way, means* or *how* of reaching this goal. Netflix recently broke with prevailing recommendation practices to embrace the *how* with the artwork personalization project, in which they personalize how recommendation is presented to their members. Thumbnail imagery for each viewer is personalized based on that person's viewing habits, instead of using the most enticing image for the biggest number of users. This has delivered improvements in customer engagement metrics, resulting in overall increase in binge-worthiness [35].

This study proposes that the *how* instead of just the end goal be personalized to each user via messages embedded with influence principles of persuasion. Adapting the *how* to individual users is advocated throughout many fields that study persuasion. In this study, we emphasize the personalizing of marketing messages to individual users based on their susceptibility to distinct persuasion principles. To empirically learn a user's susceptibility, we structure the case as a sequential decision-making problem, and propose a solution using a generalized reinforcement learning (RL) algorithm. A RL agent nominates a candidate principle to be embedded in a future message from a catalog of persuasion principles. The recommended principle is embedded in the next campaign message to the users. Message personalization and delivery is by email.

Email remains the channel of choice for customer acquisition and retention. There are over 4 billion email users worldwide, sending and receiving 3.4 billion emails per day. The number of email users is still growing, with a projected growth of 3% per year [6]. 95% percent of consumers

check their emails daily. 86% of professionals prefer to use email when communicating for business purposes. 73% of consumers named email in their top two (out of eight) in terms of preference. Stacked up against social media, 83% of consumers regard email as their preferred methods of brand communication, while one in five consumers read every email newsletter just to see if an offer is included [5, 6, 8, 9].

Email presents a unique opportunity to study user engagement given the possibility that messages could be embedded with social influence principles of persuasion and each user's response observed and measured. In the context of this study, user engagement is formalized as a sequential decision-making problem, and modelled as a Markov Decision Process (MDP). MDPs formally describe an environment for reinforcement learning. They provide a mathematical framework for modeling decision making in situations where the outcomes are partly random and partly under the control of a decision maker. The core problem of MDPs is to find a "policy" that maximizes some notion of reward. In the reinforcement learning model, an agent faces a problem of sequential decision making under uncertainty, by repeatedly interacting with an environment with unknown dynamics and receiving the rewards as shown in Figure 4.2.

At each time step, the agent selects an action and receives a reward, and the state of the system (or environment) evolves. To maximize the cumulative reward, the agent must learn the optimal behavior, reward functions and system dynamics using the observed rewards and state transitions.

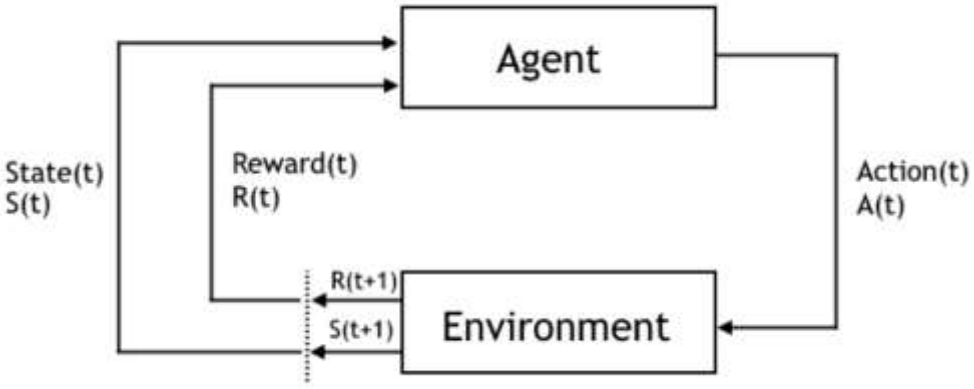

Figure 9: The reinforcement learning model (Source: Wang & Chen et al., 2017)

This must necessarily involve a trade-off between exploration and exploitation. On the one hand, it must explore different actions in various states to maintain a precise enough model of the system; on the other hand, the chosen action in a given state should be consistent with its past experience to maximize the reward.

There is no model for how the users choose to engage. They may exhibit unpredictably complex and dynamic engagement or purchasing behaviors. To determine beforehand which persuasion principle would work best for each user at a given time is a daunting and challenging task. To overcome the challenge, we propose to use a RL-based algorithm.

The goal of RL is for the agent to learn an optimal policy that maximizes the reward function or other user-provided reinforcement signal that accumulates from the immediate rewards. RL methods have been shown to be effective on a large set of simulated environments [126, 159, 105], but application to real-world problems is only now picking up speed. The two most popular classes of RL algorithms are Q- Learning and Policy Iteration. Q-learning is a type of value iteration method aimed at approximating the Q function, while Policy Iteration is a method to directly optimize in the action space. Mnih et al. proposed a Deep Q-learning Network (DQN),

which is one of the most representative algorithms in the family of value-based deep RL methods. While RL agents have achieved remarkable success in a variety of domains [188], the algorithms typically require large training data sets for convergence, making them mostly inapplicable for our application. This challenge is known as RL generalization in the literature [38, 187]. A generalized RL algorithm can train with limited data, thus achieving both sample efficiency and performance in its operation. Moreover, generalized RL algorithms can be applied in slightly different environments without any need to train the agent from scratch. In contrast, human agents have high generalized perception that performs well in unknown environments with even limited experience.

There are various studies on generalized RL. Boyan and Moore [185] proposed to select appropriate function approximate model. Another approach is to perform regularization on policy space [187]. However, none of the previous works focus on deep RL generalization. Recently Glatt, Da Silva et al. [188] proposed a framework for transfer learning in RL, but the methodology requires massive training data initially and is task dependent. The need for large amounts of data for machine learning in general, and RL in particular, makes it particularly challenging to use these methods in our setting. In a famous paper entitled, *Deep Reinforcement Learning Doesn't Work Yet*, a Google researcher has suggested that RL has so far not worked sufficiently well in solving real-world problems [79]. Other studies acknowledge that though RL has been effective on a large set of simulated environments, actual real-world deployments have been few and far between owing to several challenges including data efficiency and high dimensional state and action spaces [17], among others.

In this study, the agent does not have the benefit of being pre-trained on a large corpus of historical data. The application is quite unique in terms of objectives and largely deals with an

audience with little or no prior knowledge. Additionally, the campaigns are relatively short, aimed at small groups yielding very limited data, precluding the luxury of learning over an extended period of time. Campaigns typically last between 7 and 30 days. Initiated with Facebook ads built to target only users most likely to respond to a "lead magnet" designed specifically for health and fitness enthusiasts and hobbyists, the marketing funnel looks like Figure 10.

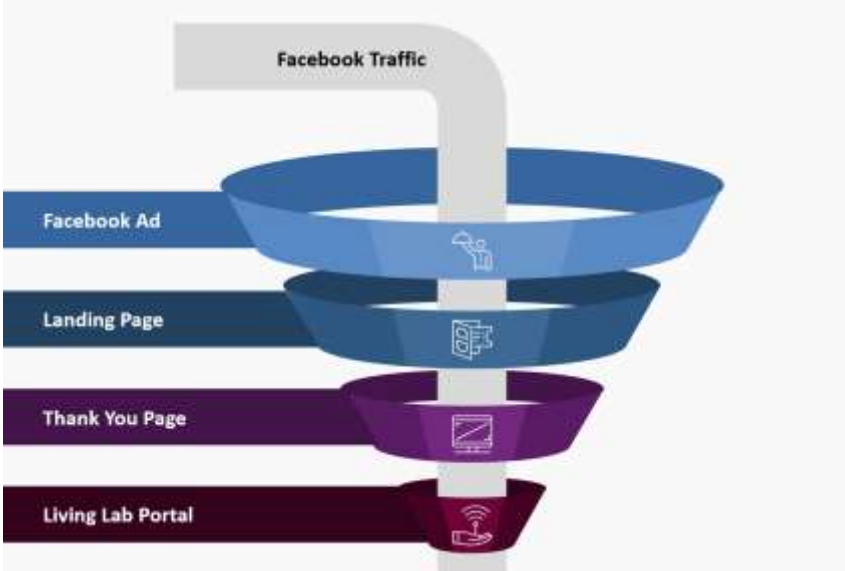

Figure 10: Lead Capture Funnel for Data-Efficient Recommender

Once opted into the subscribers' list, a welcome message with links to the promised information i.e., a free wellness report in either text or video is emailed to the user. Over the next several days, a number of email messages with embedded persuasion principles are sent to user's inbox in an attempt to learn what persuasion strategies could possibly trigger deeper levels of engagement leading to replies, reviews, ratings and purchases. Implementation of our strategy is restricted to the subject of the email messages, in line with direct marketing studies which suggest that on the average, five times as many people read the headline as read the body copy [136].

After the welcome email, the next few messages have persuasion principles embedded in the subject lines as follows:

1. *"1000s of wellness enthusiasts love the new Tiny Habits book."*
2. *"Harvard's Professor Ford recommends reading the new Tiny Habits book."*
3. *"Only 48 hours left to get your signed copy of the new Tiny Habits book."*

Where the first implements the persuasion principle of consensus, the second authority, and the third scarcity. After the initialization messages which typically lasts for about 3 days, various combinations of strategies may be deployed to boost engagement and drive participation in the Living Lab. The multiple-principle subject lines are structured as follows:

4. *"[Tiny Habits] 1000s participating. Only 48 hours left."*
5. *"[Tiny Habits] Harvard's Professor Ford recommends it, only 18 hours left."*
6. *"[Tiny Habits] Harvard's Professor Ford recommends it. 1000s are participating."*

Where #4 combines the persuasion strategies of consensus and scarcity, #5 combines authority and scarcity, while #6 is a combination of authority and social proof. Our learning agent has a severely limited window of opportunity to determine the near-optimal persuasion principle most likely to receive the desired click response from each individual user. Within succeeding 24 hours, before the next message broadcast, the agent delivers a list of recommendations matching users to persuasion principles similar to Table 2. A Python script reads this table, embeds the suggested principle for each user and prepares the messages for broadcast scheduling using a commercially available autoresponder email system such as Mailchimp, Aweber or Hubspot. After the messages are sent and delivered, user engagement clickstream is transmitted to the learning agent within a 24-hour period to initiate the recommender process for the next messaging cycle.

User	Persuasion Principle
User1	Authority
User2	Consensus
User3	Scarcity
:	:
:	:
UserN	Reciprocity

Table 2: Sample recommendation for next messaging cycle

Unlike purely data-driven recommenders, such as ecommerce and online advertisements, with virtually limitless troves of data for training and testing, the data-efficient, persuasion-based recommender in this study requires a different approach. Real-world systems are fragile and expensive enough that the data produced is costly and policy learning must necessarily be data efficient. In this setting, our offline log does not contain anywhere near the amount of data or data coverage that current algorithms expect. To address this challenge, we introduce a novel process for combining mechanistic and data-driven approaches.

Mechanistic models have served as a foundation for the study of behavior change systems for years. Such models define a socio-technical information system with behavioral outcomes designed to form, alter, or reinforce attitudes, behaviors or an act of complying without using deception or coercion [190]. However, their success has been limited by a lack of accurate inputs and precision of predictions. Data driven methods such as machine learning and deep learning, on the other hand, can overcome these limitations by examining patterns in data to produce more accurate predictions. This study proposes a means to hybridize these two methodologies by

combining user engagement data with domain knowledge, such as what stage of the customer journey to apply specific influence principles that would motivate engagement and lead to a lift in response as high as 2,400% (consensus), or as low as 45% (scarcity) [37].

As noted earlier, traditional RL algorithms are not that effective in handling state space complexity and large state spaces [17]. There are a number of avenues for reducing the dimensionality of the state space for RL application settings. In the context of email engagement campaigns, we are dealing with historical data in terms of strategies employed by prior messages (i.e., persuasive principles) and the complete user response/activity history. Without loss of generality, we propose an exponential weighted moving average scheme for compressing this history. This emerged as the only viable solution after years of attempting to segment the users into persuasion clusters, an approach which proved immensely difficult with our setting which features short campaign cycles coupled with extremely limited and sparse data samples. This novel approach to state space representation obviates the need for copious amounts of data for training the RL algorithms.

Overall, pairing a limited dataset with universal influence principles known to motivate people of all backgrounds and cultures to take some form of action, both online and offline, has proven indispensable in our setting where the agent needs to learn quickly, often in four interactions or less. Campaigns are of short duration, sometimes targeting brand-new prospects acquired through a wide variety of traffic sources. Each campaign is different; however, the core problem remains the same: the search for an optimal method of engagement, given the need to make decisions – every day, week or month – requiring repeated engagements until the campaign ends, with the user purchasing a product, or not engaging at all, or in the worst-case scenario, unsubscribing from the list.

The rest of this chapter is organized as follows: Section 2 describes the experimental setting, Section 3 proposes the novel Deep RL algorithm which can be trained with a limited dataset, Section 4 deals with learning and measuring the performance of proposed algorithm on a simulated dataset related to an email marketing campaign, and Section 6 discusses the results from a real-world case study used to demonstrate the performance of the developed algorithm and compare the results with a human expert and a baseline control. Finally, the research is summarized in Section 7.

4.2 The Setting

4.2.1 Building a Data-Efficient Recommender

High-performance recommenders often considered the engine driving deep user engagement for leading digital platforms have a major weakness. They share a common vulnerability that, despite their strength, could lead to their eventual downfall. They are all data hungry. In a world with near-infinite data coupled with lax data privacy rules this might not pose such a problem. However, as noted recently by Tim Cook, CEO of Apple, the first publicly traded U.S. company to hit $2 trillion in market value[1]:

> *"Technology does not need vast troves of personal data stitched together across dozens of websites and apps in order to succeed. Advertising existed and thrived for decades without it. And we're here today because the path of least resistance is rarely the path of wisdom.*
>
> *"At a moment of rampant disinformation and conspiracy theories juiced by algorithms, we can no longer turn a blind eye to a theory of technology that says all engagement is good engagement, the longer the better. And all with the goal of collecting as much data as possible. It is long past the time to stop pretending that this approach doesn't come with a cost."*
>
> Tim Cook, CEO, Apple
> January 28, 2021, Data Privacy Day Speech[2]

[1] https://www.forbes.com/sites/sergeiklebnikov/2020/08/19/apple-becomes-first-us-company-worth-more-than-2-trillion
[2] https://www.businessinsider.com/tim-cook-takes-swipe-at-facebook-social-catastrophe-2021-1

With our emphasis on data efficiency, this chapter demonstrates one proof of concept that might become more relevant as we envisage a future with stronger data privacy rules and regulations.

4.2.2. Dataset

We conduct an experiment on a mix of sampled and simulated offline dataset collected from the email server of a commercial digital publishing site and deploy our system online for two successive marketing campaigns of 7 days each. The recommendation algorithm will make a recommendation every 24 hours after an email message has been sent and user response recorded (click or no click). The subject line of the email message incorporates a distinct principle of persuasion (or a combination) selected from the six universal principles of persuasion [30]. The system is initialized with principles known to have the highest probability of eliciting a response (e.g., click-open, click-through, click-engage, click-purchase). The basic statistics for the sampled data are shown in Table 3. In the offline stage, the training data and testing data are separated by time order (the last two weeks are used for testing). During the online deployment stage, we use the offline data to pre-train the model.

To capture and analyze information of state and action over time, and to smooth out short-term fluctuations and highlight longer-term click trends, we design a state space transformation based on exponential weighted moving averages (EMA), a first-order filter that applies weighting factors which decrease exponentially. The weighting for each older data point decreases exponentially, never reaching zero.

Stage	Duration	# of users	# of campaigns
Offline (training & testing)	6 weeks	25	40
Online	2 weeks	150	2

Table 3: Statistics of the sampled dataset

For better illustration, we report below the raw click data for a sample user1 (see Table 4). At the end of day 1, we have one day history (last row). At the end of day 2, we will have two days of history (last two rows). This growing history presents a challenge; the dimensionality of the state space is increasing with each campaign day. To address this problem, without loss of generality, we recommend that the history associated with each state space variable (i.e., each column of Table 4) be compressed at the end of each campaign day using an EMA scheme with an appropriately selected discount rate (η). This ensures that the cardinality of the state space remains the same for every day of the campaign, reducing the needs for large datasets for training the RL agents. The result from the one-dimensional EMA scheme is reported in Table 5, which form the input into the RL neural network.

P1	P2	P3	P4	P5	P6	P7	P8	P9	Open	Click	Engage	Buy	Unsub	Day
0	0	1	0	0	0	0	0	0	1	1	1	0	0	7
1	0	0	0	0	0	0	0	0	1	1	0	0	0	6
0	0	0	0	0	0	0	1	0	1	0	0	0	0	5
1	0	0	0	0	0	0	0	0	1	1	0	0	0	4
1	0	0	0	0	0	0	0	0	1	0	1	0	0	3
0	0	0	0	0	0	1	0	0	1	1	1	0	0	2
0	0	0	0	0	0	1	0	0	1	1	1	0	0	1

Table 4: Raw clickstream data for single user over 7-day campaign

P1	P2	P3	P4	P5	P6	P7	P8	P9	open	Click	Engage	Buy	Unsub	Day
0	0	1	0	0	0	0	0	0	1	1	1	0	0	7
0.15	0	0.85	0	0	0	0	0	0	1	1	0.85	0	0	6
0.1275	0	0.7225	0	0	0	0	0.15	0	1	0.85	0.7225	0	0	5
0.258375	0	0.614125	0	0	0	0	0.1275	0	1	0.8725	0.764125	0	0	4
0.3696	0	0.522	0	0	0	0	0.108	0	1	0.7416	0.79950	0	0	3
0.3141	0	0.4437	0	0	0	0.15	0.0921	0	1	0.7803	0.8295	0	0	2
0.2670	0	0.3771	0	0	0	0.27	0.0783	0	1	0.8133	0.8551	0	0	1

Table 5: Transformed EMA version of the same user data ($\eta = 0.5$)

The RL in this application is a deep neural network which, in the literature, typically consist of a set of input units, multiple hidden layers containing hidden units (also known as nodes or neurons), and a set of output unit, with connections running between those nodes as depicted in Figure 11.

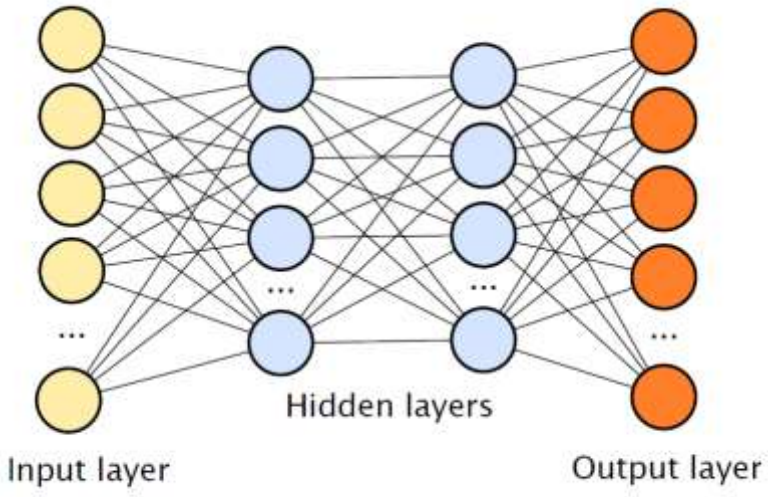

Figure 11: A Deep Neural Network Architecture (Source: Marcus, G., 2018)

4.3 Problem Definition

We define our problem as follows:

When a new campaign email is about to be sent, our RL agent is going to recommend a persuasion principle to be included in the message. The user receives that email message and responds in the form of clicks. These users and their responses constitute the environment while the algorithm is tasked with nominating a candidate action from a catalog of persuasion principles to be embedded in the next message to the user. After each day's message (t denotes time), the agent selects the principle for the next message, a_t, and receives a reward from the user r_t in form of open, click, engagement and purchase. For every campaign day we have $X_t = (a_t, r_t)$. Given X_t, the agent needs to recommend the principle for the next day's message with the goal of maximizing long term user responses in the form of more clicks, which, in this study, act as a proxy for higher engagement.

4.3.1 Markov Decision Process Formulation

We formulate the persuasion principle selection and recommendation problem as a Markov decision process (MDP), defined as a tuple $\zeta := \{S, A, P, R, S_0\}$, where S is the state space, A is the action space, the transition probability function is defined as $P(s, a, s') = \Pr(s_{s+1} = s' \mid s_t = s, a_t = a), a \in A; s, s' \in S$, $R: S \times A \to \mathbb{R}$ is a reward function and S_0 is the initial state distribution. A policy $\pi = S \to A$ is a mapping from state space to action space. At a given time-step t, the agent draws an action from the policy. The agent then receives a reward $r_t = R(s_t, a_t)$ and transitions to the next state s_{t+1} according to the transition probability. This process produces a trajectory $\tau = \{s_0, a_0, r_0, s_1, a_1, r_1, ..., s_T, a_T, r_T\}$. The value function at state s given a discount factor $\gamma = [0,1]$ is defined by the expected discounted return under policy π as $V_\gamma^{\pi}(s) := \mathrm{E}_\tau(\sum_{t=0}^{T} \gamma^t r_t \mid s_0 = s)$. In our system $< S, A, P >$ are defined as follow:

Action A is a finite set of persuasion principle-based messages which include reciprocity, scarcity, scarcity, consensus, authority, consensus & scarcity, and consistency.

State S in our system represents information about open, click, engagement, purchase of users at each campaign day, historical actions and number of days remains in the campaign. Since the environment is very sparse to prevent overfitting, we propose the use of exponential weighted moving average (EMA) as state transformer. EMA is a first-order infinite impulse response filter that applies weighting factors which decrease exponentially. The weighting for each older data decreases exponentially, never reaching zero. The transformation contributes to stability and coverage of RL. Formally, state representation of t^{th} day of campaign is as follows

$$s_t = [EMA(k_t), EMA(a_{t-1}), t^*] \quad (1)$$

where t^* is number of days remains in the campaign, $EMA(k_t)$ and $EMA(a_t)$ are defined as follows:

$$EMA(k_t) = \begin{cases} k_t, & t = 1 \\ \eta(k_t) + (1-\eta)EMA(k_{t-1}), & t > 1 \end{cases} \quad (2)$$

where $\eta \in [0,1]$ is a constant smoothing/discounting factor, t is number of days passed from the beginning of campaign and k_t is a vector consist of number of open, number of click, engagement indicator and purchasing indicator at t^{th} day of campaign

Similarly, the EMA of action is represented as follows:

$$EMA(a_t) = \begin{cases} a_t, & t = 1 \\ \eta(a_t) + (1-\eta)EMA(a_{t-1}), & t > 1 \end{cases} \quad (3)$$

Transition P is the probability of observing s' at time t after observing state s and taking action a at time t. In this case, the uncertainty comes from user's response to the received email message.

4.3.2 The Learning Algorithm

The expected reward is represented with a Q-function (action-value function) given state s and action a is defined as

$$Q_{\gamma}^{\pi}(s,a) = E_{\tau}(\sum_{t=0}^{T} \gamma^{t} r_{t} \mid s_{0} = s, a_{0} = a) \quad (4)$$

To compute the Q-function efficiently, function approximation methods are widely used approaches in literature. DQN approximates the action value function through a deep neural network (Figure 12). The input of the network takes the current state s_t and outputs $|A|$ corresponding to the state-action values of each persuasive principles. However, the method requires many episodes and much data to train the deep neural network and convergence could take significantly longer.

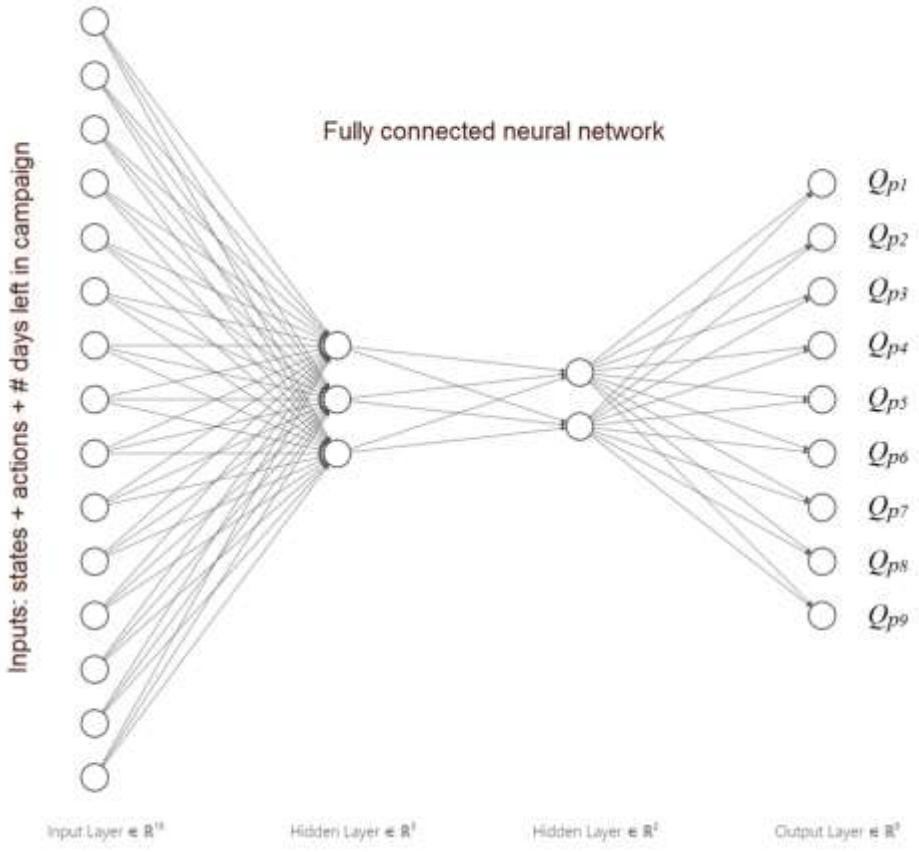

Figure 12: Schematic illustration of the DQN

The inputs to the DQN consist of 5 states, 9 actions, and 1 variable tracking days left in the campaign, followed by two fully connected layers with outputs corresponding to the state-action values of persuasion principles Q_{p1} to Q_{p9}

In our study, the agent can only interact with the environment in a limited way. Hence the trajectory length T and number of generated trajectories are limited. Furthermore, in our setting, environment's data are limited. Since the data are not truly representative of the actual state space, DQN suffers from high estimation variance. To reduce the variance and accommodate the sparse environment's data, the time horizon in Q approximation function must be limited in order to

lower the Q function approximation's values. To impose the constraint on the special environment setting, we propose a regularized DQN.

Specifically, we added $\lambda Q^2_{\boldsymbol{\theta}}(\varphi(s,a),a), \lambda > 0$ to TD error for training the DQN. The penalty term assures that Q does not become a large value. As a matter of fact, with a fixed γ and r, a reduced Q value considers a short trajectory length which is efficient in the case where only limited interaction with environment is possible. The proposed algorithm containing the regularized DQN and transformation function, EMA of state and action, is outlined as follows:

4.3.3 The Algorithm

Initialize $\alpha \in [0,1]$ learning rate, ε epsilon value, β epsilon decay rate, N number of iteration, B batch size , number of episodes M , D replay buffer capacity, T iteration, η exponential moving average parameters.

Algorithm 1

Initialize $\alpha \in [0,1]$ learning rate, ε epsilon value, β epsilon decay rate, N number of iteration, B batch size , number of episodes M , D replay buffer capacity, T iteration, η exponential moving average parameters

Initialize action-value function Q with weights $\boldsymbol{\theta}$

Initialize target action-value function Q with weights $\boldsymbol{\theta}' = \boldsymbol{\theta}$

Interact with a random policy to fill the reply buffer

For $t = 1, T$ do

Select an action a_t with probability of ε

Otherwise select $a_t = \arg\max_a Q_\theta(\varphi(s_t), a)$ given $\varphi(s_t) = [EMA(s_t), EMA(a_{t-1}), t]$

Obtain $(s_t, a_t, s_t^{'}, r)$ from environment and store in replay buffer

Sample a batch of $(s_j, a_j, s_j^{'}, r_j)$ $j = 1..., B$ from reply buffer

calculate $\varphi(s_j) = [EMA(s_j), EMA(a_j), a_j, t]$ for every sample in batch

$y_i = r_j + \gamma \max_{a'} Q_{\theta'}(\varphi(s'_j), a')$ if s'_j is not a terminate state; $y_i = r_j$ otherwise

Update $\boldsymbol{\theta}$ according to the TD error

$$L = E_{(s_j, a_j, s_j^{'})}\left([\sum_{j=1}^{B} Q_{\boldsymbol{\theta}}(\varphi(s_j), a_j) - y_j]^2 + \lambda \sum_{j=1}^{B} Q^2{}_{\boldsymbol{\theta}}(\varphi(s_j), a_j) \right), \lambda > 0 \quad (5)$$

Every N step copy weights from $Q_{\boldsymbol{\theta}}$ to $Q_{\boldsymbol{\theta}'}$

Theorem 1 shows that to make traditional DQN efficient to handle sparse environment's data, the discount factor must be reduced.

Theorem 1

Given reward $r \geq 0$, discount factor γ, tuning parameter $\lambda > 0$ and learning rate α, the TD error in proposed regularized DQN (algorithm1) is equivalent to standard TD in traditional DQN with $\gamma^{'}$ discount factor where $\gamma^{'}$ is formulated as follows:

$$\gamma' = \gamma - (\lambda / \alpha) \frac{r + \sum_{j=1}^{B} \max Q_{\theta}(s_j', a_j')}{\sum_{j=1}^{B} \max Q_{\theta'}(s_j', a_j')}$$

Note that $(\lambda / \alpha) \frac{r + \sum_{j=1}^{B} \max Q_{\theta}(s_j', a_j')}{\sum_{j=1}^{B} \max Q_{\theta'}(s_j', a_j')} \geq 0$, thus reducing the discount factor.

This is in line with recent studies [7].

4.4.5 Offline Campaign

For the offline experiment, we collected historical data on 15 active users who had participated in a short email marketing campaign in the past 30 days. Selected users shared other similar features such as responding to the same Facebook ad, joining the subscriber list through the same opt-in form, and demonstrating some level of engagement through specific actions such as email opens, click-throughs, and replies. Clickstream data over a prior 7-day campaign period were collected from the email marketing system of an online bookstore. From this real-world sample, more data were simulated to cover a total of 40 campaigns of 7-day durations each for each user. Table 6 illustrates a typical 7-day clickstream for a single active user.

Days	Persuasion principles	User response
Day 1	Reciprocity	Open, Click Engage
Day 2	Scarcity	Open, Click, Engage
Day 3	Scarcity	Open, Click, Engage
Day 4	Consensus	Open, Click
Day 5	Authority	Open, Click, Engage
Day 6	Consensus+ Scarcity	Open, Click, Engage, Buy
Day 7	Consistency	Open, Engage

Table 6: Sample states and actions for 7-day campaign

As shown in the table, possible user's responses include *Open*, *Click*, *Engage* and *Buy*. *Engage* is defined as a binary value, which is 1 when a user opens or clicks the link in the email message more than once and 0 otherwise. The reward structure is such that the agent receives a higher reward when the user engages with the message or buys a product, while receiving a lower

reward if the user fails to engage or unsubscribes. We consider episodes with length of 7 days and train the DQN agent with 10,000 iterations. Given that the experiment is conducted on an offline dataset, we have limited access to the users' response s' given action a and current state s. To overcome this limitation, a distance function is used to select next state s'. First, we structure the available data into a tuple of $(s_i, a_i, s_i', r_i),\ i = 1.., N$. Second, we randomly choose s' from $(s_j, a_j, s_j'),\ j = 1,.., n$ that satisfies the following inequality:

$$\| (s, a) - (s_j, a_j) \| < \theta \tag{6}$$

where θ is a constant between 0 and 1. If none of the tuples $(s_i, a_i, s_i', r_i),\ i = 1.., N$ satisfies (6), we end the episode and reset the environment. Finally, the hyper-parameters of the proposed DQN are tuned using a grid search algorithm. We observe the discount factor, batch size, learning rate and buffer size have the most impact on the performance of the algorithm.

The best tuning parameters obtained by grid search are batch size = 5000, gamma = 0.09, replay buffer = 2000 and learning rate = 0.00005. Previous experimental results have shown that experience replay vastly improve the learning performance, partly because it can reduce the correlation between the samples (Chen et al., 2018). Furthermore, learning rate controls how quickly the model adapts to the problem and data. To confirm the performance of the DQN agent, we present 95% confidence interval of collected rewards by the agent over 3 runs. Each point of the plot is calculated by the formula:

$$\overline{X} = Z \frac{S}{\sqrt{n}}$$

Where n is number of runs and S is empirical standard deviation. As the figure shows the confidence interval, the shaded area, of rewards after 4000 iteration is narrow illustrating the algorithm's results have lower variability.

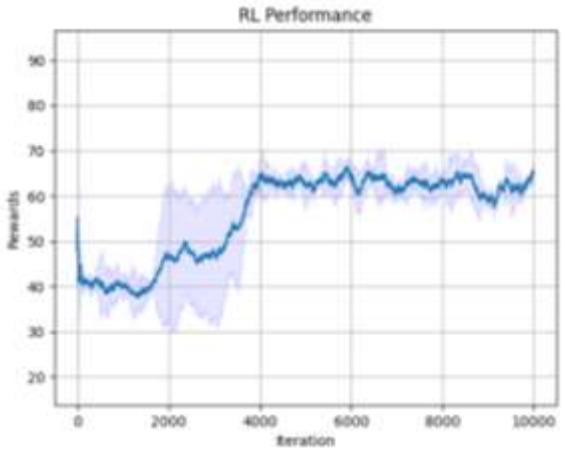

Figure 13: Agent performance at 95% confidence interval

We validate the performance of our proposed DQN agent by comparing the collected reward of the agent with a random agent which takes random action independent of current state and next observed state. Several hyper-parameters were tuned. We observed the discount factor and batch size have the most impact on the algorithm performance. Figure 4 shows the rewards for different parameters setting. By increasing the batch size and reducing the discount factor performance of DQN improves (14e, 14f, 14g). But increasing the batch size without lowering the discount γ does not improve performance (14a, 14b, 14c). High discount factor or low batch size degrade learning stability and lead to lower cumulative reward. The results match several studies demonstrating that lower discount can significantly improve generalization performance when learning from limited data.

Bertsekas & Tsitsiklis (1996) show that lower γ increases convergence rate in many RL algorithms, but other works suggest that it can also improve final performance in the case of limited data. Jiang et al. (2015) studied a model based RL setting and suggested that in the limited data regime, the performance can be improved by using a low discount factor in the planning phase. Amit & Meir (2020) show that in the learning setting, lowered discounts allow for better generalization.

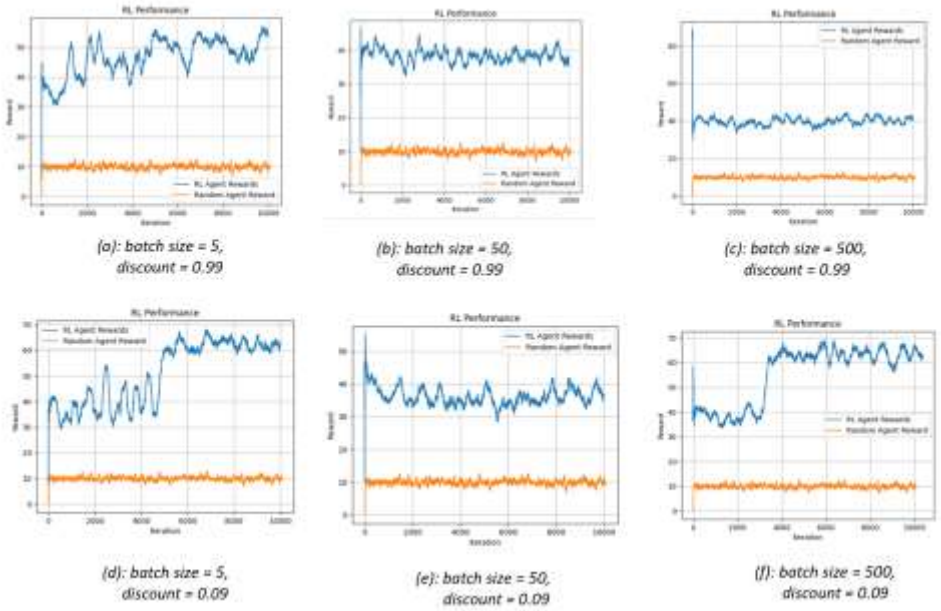

Figure 14: Hyper-parameter tuning

Batch size increase without lowering discount γ degrades performance *(a,b,c);*
Batch size increase with lower discount factor γ improves performance *(d,e,f)*

4.5 Online Campaign

To measure the performance of our algorithm and compare it with the human and random agents, we deployed the proposed methodology to a live email marketing campaign. We identified 150 potential users for the campaign and randomly assigned them to RL, human and control groups with 50 members each. For the purpose of initialization, we send the same persuasion principles to all 150 users for the first 3 days and record each user's click responses. The trained DQN agent is deployed on the RL group to recommend the persuasion principle to be employed in the message for each user starting the 4^{th} day of the campaign . An expert with domain knowledge (Human agent) selects the principles to be sent to users in human group. Finally, the control group receives randomly selected principles (Control agent).

The treatment continues for two consecutive campaigns of 7 days each. In summary, all users receive same treatment for the first 3 days of the first campaign. The action and response history from the first 3 days are used to initialize the state space for the RL agent. The choice of principles used for the initialization is based partly on the influence literature and on the observed click behavior of similar users who respond to social media ads on Facebook. A key assumption made is that users acquired through the same traffic source using the same ads, landing pages, and "thank you" pages will exhibit similar engagement behavior as they are more likely to share the same objectives, interests and preferences. For the remaining days of the campaign, this assumption is put to the test as the agent relies solely on the offline policy to recommend the principles most likely to elicit a response from each user within the RL group. Though the agent was pre-trained on offline data associated with a different set of users, it seems reasonable and appropriate to assume that transfer learning is possible given the similarity of user populations.

Table 7 shows the number of daily opens, clicks, and engages of RL, Human and Control agents, including the first 3 initialization days. The table clearly demonstrates that the RL agent outperforms in daily number of opens, clicks and engagements.

		Days	RL Open/Human Open/ Ctr Open/	RL Click/Human Click/ Ctr Click	RL Eng/Human Eng/ Ctr Eng
1st Campaign	Initialization Phase	1	34 / **36** / 28	**23** / 20 / 12	12 / **14** / 9
		2	**30** / 27 / 24	**16** / 11 / 9	**10** / 6 / 8
		3	**26** / 21 / 25	11 / 13 / **14**	**8** / 3 / 7
	RL Policy based Recommendations	4	**39** / 15/ 14	**24** / 9 / 5	**34** / 0 / 3
		5	**34** / 14 / 30	**24** / 6 / 9	**22** / 5 / 12
		6	**45** / 16/ 12	**25** / 14 / 4	**21** / 6 / 3
		7	**25**/ 14 / 11	**21**/ 3 / 5	**25** / 6 / 3
2nd Campaign		8	**26** / 12/ 9	**11** / 1 / 5	**9** / 1 / 1
		9	**35** / 7 / 7	**11** / 2/ 5	**18** / 2 / 1
		10	**42** / 17 / 5	**14** / 7 / 1	**12** / 7 / 0
		11	**33** / 9 / 9	**11/** 2 / 4	**18** / 3 / 2
		12	**56** / 8 / 8	**43** / 3 / 1	**20** / 3 / 4
		13	**84** / 17 / 16	**31** / 8 / 3	**26** / 7 / 6
		14	**45** / 21 / 13	**14** / 7 / 5	**15** / 3 / 2

Table 7: Statistics of User Reponses under RL, Human and Control Agents

We define rewards as average number of open, click and engage. Figure 15: illustrates the rewards attained by RL, Human and Control agents. RL agent achieves highest reward for each single day throughout the 2 live campaigns. We observe that Human agent is slightly better than Control agent on certain days.

Figure 15: Live Campaign Rewards under RL, Human and Control Agents

Figure 1615 shows accumulated number of opens, clicks and engages of the three agents throughout the duration of the two live campaigns. The data shows that Human and Control agents draw similar level in respect of engagement and RL agent outperforms the other agents.

4.6 Results

In this study, email messaging is used as the vehicle to deliver persuasion principles to the user. At a time of declining click-through rates with marketing emails, business executives continue to show more interest in the email channel owing to higher-than-usual return on investments compared to other channels. To increase engagement with email messages, we combine the psychological principles of persuasion with the data-driven methods of RL. A novel regularized DQN able to train and perform well using limited historical data is proposed. Furthermore, we design a state-space dimensionality reduction based on EMA. With this transformation, the regularized DQN can capture and analyze information of state and action over time. A simulation and a real live campaign are implemented to verify the proposed methodology and to demonstrate its superior performance compared to a human expert and a control baseline.

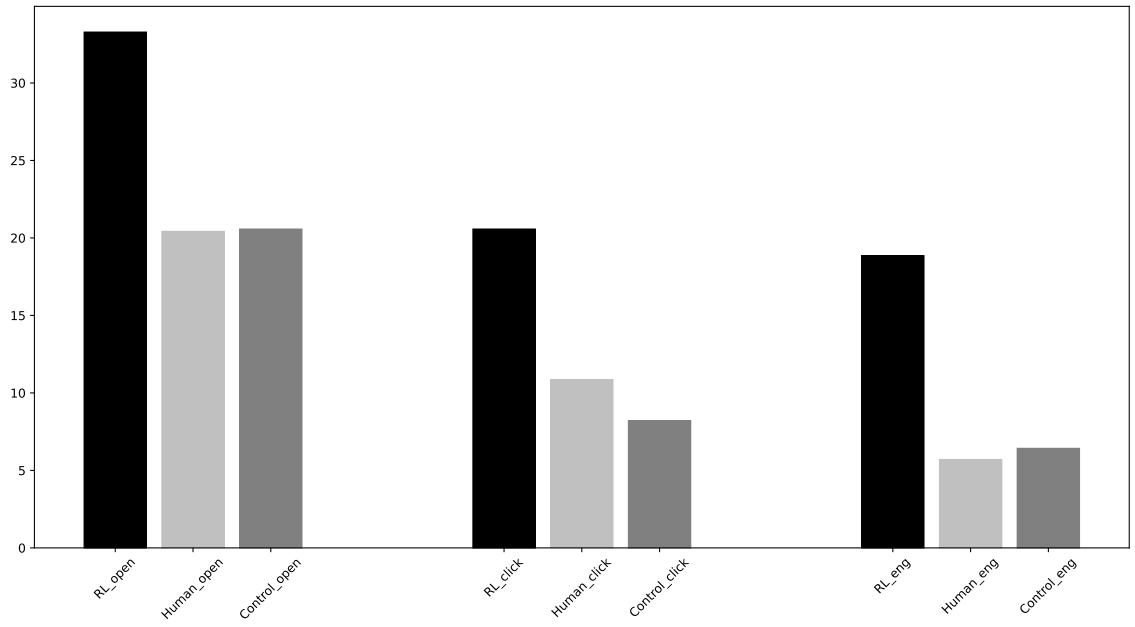

Figure 16: Accumulated Engagement Statistics

4.7 Discussion

We attempted to frame user engagement as a sequential decision-making problem, modelled as MDP, and solved using a generalized RL algorithm. To the best of our knowledge, this work is the first attempt to define user engagement as a RL recommender problem, combining persuasion principles and data-driven methods, to build a sample-efficient recommendation agent to nominate candidates from a catalog of persuasion principles most likely to drive higher engagement. For operationalization of this recommender, more detailed investigation of predictive performance with a diversity of audiences should be conducted. This work shows that in our study setting – online marketing – user clicks and other responses present significant behavioral clues which can be used to target to the right influence principle to the right user for enhanced engagement. As more data becomes available, the prediction and recommendation power should

simultaneously improve. However, this might not always be the case, given the fact that "thin slices" of expressive behavior sampled from the behavioral stream can sometimes approach the level of accuracy usually associated with analyzing massive amounts of data [6].

4.8 Limitations

One limitation associated with the current approach is the total dependence on offline learning from history associated with a different user dataset. Unless the user populations exhibit close similarities in terms of traffic sources, industry niches, purchase histories and previous offers presented, transfer learning might not be effective. With any other setting such as a longer campaign horizon, learning could be adversely affected. If the agent succeeds in learning at all, consistent results may not be guaranteed. Also, there are limitations with the validation set in terms of size (one company, one campaign) and in terms of diversity (one industry). The training sample is limited with a disproportionately large amount of the dataset partially simulated. A more diverse data sample, coupled with some online learning, could improve the results, however, these could potentially compromise repeatability and consistency of predictions.

4.9 Conclusion

This work builds on a growing academic and practitioner attention to the need for increased user engagement with email messages in particular, and online messages in general. Persuasion-based recommender system that combines domain behavioral knowledge with data-driven RL methodology can help with effectively suggesting the right persuasion principles most likely to elicit the desired response from the right customer in future messaging campaigns. The RL system in this study reveals how persuasion principles can lead to more clicks from each user and lead to higher engagement both at the individual and group level. Future studies could incorporate more

advanced algorithms and modelling techniques such as few-shot learning to boost user engagement at a tiny fraction of the today's data and infrastructure costs.

Proof of Theorem:

$$[\sum_{j=1}^{B} Q_{\boldsymbol{\theta}}(s_j, a_j) - y_j]^2 + \lambda \sum_{j=1}^{B} Q^2{}_{\boldsymbol{\theta}}(s,a) = [\sum_{j=1}^{B} Q_{\boldsymbol{\theta}}(s_j, a_j) - r_j + \gamma \max_{a'} Q_{\boldsymbol{\theta}'}(s'_j, a')]^2 + \lambda \sum_{j=1}^{B} Q^2{}_{\boldsymbol{\theta}}(s,a) \quad (7)$$

For convention, we consider the following substitutions:

$$Q'^2{}_{\boldsymbol{\theta}}(s,a) := \sum_{j=1}^{B} Q^2{}_{\boldsymbol{\theta}}(s,a), \quad Q'_{\boldsymbol{\theta}}(s,a) = \sum_{j=1}^{B} Q_{\boldsymbol{\theta}}(s,a), \; Q'_{\boldsymbol{\theta}'}(s,a) = \sum_{j=1}^{B} Q_{\boldsymbol{\theta}'}(s,a) \quad (8)$$

According to gradient descent of proposed loss function (5) in regularized DQN, we have the following parameter update rule:

$$\begin{aligned} &\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + \alpha[-\mathbf{r} - \gamma \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}', \mathbf{s}') + Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s})].\nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) + \lambda Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) = \\ &\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \alpha \mathbf{r} \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) - \alpha\gamma \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}', \mathbf{s}') \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) + \alpha Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) + \lambda Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) \end{aligned} \quad (9)$$

where $i = 1, ..., n$ is gradient decent iteration.

Gradient descent of traditional DQN's loss function gives the following update rule:

$$\begin{aligned} &\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i + \alpha[-\mathbf{r} - \gamma' \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}', \mathbf{s}') + Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s})].\nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) = \\ &\boldsymbol{\theta}_{i+1} = \boldsymbol{\theta}_i - \alpha \mathbf{r} \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) - \alpha\gamma' \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}', \mathbf{s}') \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) + \alpha Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) \end{aligned} \quad (10)$$

With the assumption that a constant discount factor $\gamma' \neq \gamma$, the traditional DQN and our proposed algorithm produce same Q value and neural network's parameters, we derive the discount factor in terms of the regularized DQN discount factor. By subtracting (9) and (10) the following equations hold:

$$\begin{aligned} &\alpha\gamma' \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}', \mathbf{s}') \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) - \alpha\gamma \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}', \mathbf{s}') \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) + \lambda Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) = 0 \\ &\alpha \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}', \mathbf{s}') \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s})(\gamma' - \gamma) = -\lambda Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) \nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a}, \mathbf{s}) \end{aligned} \quad (11)$$

Using (11), we can derive the discount factor of traditional DQN γ' as term of γ

$$(\gamma' - \gamma) = \frac{-\lambda Q_{\boldsymbol{\theta}_i}(\mathbf{a},\mathbf{s})\nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a},\mathbf{s})}{\alpha \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}',\mathbf{s}')\nabla Q_{\boldsymbol{\theta}_i}(\mathbf{a},\mathbf{s})} = \frac{-\lambda Q_{\boldsymbol{\theta}_i}(\mathbf{a},\mathbf{s})}{\alpha \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}',\mathbf{s}')}$$
$$\gamma' = \gamma - \frac{\lambda Q_{\boldsymbol{\theta}_i}(\mathbf{a},\mathbf{s})}{\alpha \max_{\mathbf{a}'} Q_{\boldsymbol{\theta}'}(\mathbf{a}',\mathbf{s}')} \tag{12}$$

By reverting the substitution in (8), and definition of Q function, the following equation is derived:

$$\gamma' = \gamma - (\lambda/\alpha)\frac{r + \sum_{j=1}^{B} \max_{a_j'} Q_{\theta}(s_j', a_j')}{\sum_{j=1}^{B} \max_{a_j'} Q_{\theta'}(s_j', a_j')} \tag{13}$$

CHAPTER 5: CONCLUSIONS AND FUTURE RESEARCH

This research contributes to and advances the literature on maximizing user engagement using the reinforcement learning approach. While machine learning methods have been used to build recommender systems aimed at boosting user engagement, there is a heavy price to pay in terms of the amount of data that current algorithms expect, as well as the issue of high dimensional state and action spaces associated with practical, real world problems. For leading technology firms, user engagement is now the engine driving online business growth. Many companies have pay incentives tied to engagement and growth metrics. Even as recommender systems algorithms have emerged as the tool of choice in the business of maximizing engagement, the problem of data and sample inefficiency continues to worsen. The setting of this study makes such algorithms inapplicable.

We introduce a novel approach that eliminates or at least greatly reduces the need for copious amounts of training data, requiring a deviation from a purely data-driven approach. By incorporating domain knowledge from the literature on persuasion into the message composition, we successfully train the reinforcement learning (RL) agent in a sample efficient and operant manner. In our methodology, the RL agent nominates a candidate from a catalog of persuasion principles to drive higher user response and engagement. To enable the effective use of RL in our setting, we first build a reduced state space representation by compressing the data using an exponential moving average scheme. Then a regularized DQN agent is deployed to learn a behavior policy, which is then applied in recommending one (or a combination) of six universal principles most likely to trigger a response from each individual user during the next message cycle. We now present a summary of our research along with contributions and discuss future directions of research that may result from our work.

5.1 Summary

5.1.1 Heuristic Application of Persuasion Principles to Increase User Engagement

We explored the use of one or more principles of persuasion in email messages broadcast to subscribers of an online publishing website, with the goal of increasing several dimensions of user engagement. Unlike the general belief that increasing engagement can be achieved only through the deployment of sophisticated algorithms, we showed that simple treatments like adding a principle of persuasion to the subject line of email messages can have substantial impact on user engagement. We incorporated more than one principle in one case and continued with the message broadcast over an extended period. The selection of persuasion principles was done heuristically, using domain knowledge and as such limited. The proposed method enabled us to increase user engagement in the treatment group by up to 100% over the baseline. The key assumption is that users are predisposed to respond to certain persuasion principles and embedding these principles in the message header or body copy will lead to higher response.

If persuasion principles are added to messages, overall user engagement will increase, but the reliability of this assertion is questionable since we are only dealing with average effects over a group of users and are not yet able to determine the effect of specific principles on specific users. While this correlation between persuasion and response cannot be precisely measured by our simple heuristic approach, it nevertheless provides a pointer to the potential impact that persuasion principles can have on boosting deep user engagement signals. In the ideal case where the principles are precisely targeted to individual users, the proposed strategy can lead to a substantial increase in user engagement by a factor of 3 to 10. In this heuristic case, based wholly on domain knowledge, where selection of persuasion principle is imperfect, the response and engagement tend to improve with the right selection, but degrades if the wrong principle is applied. The

degradation persists as more of the wrong principles are selected, resulting in overall lower engagement. On the contrary, given better choices, the decline can be slowed, stopped, and ultimately reversed. Therefore, improvement in domain knowledge and audience behavior can become vitally important in the quest to increase user engagement in the absence of rigorous algorithmic processes.

Finally, we demonstrate that the addition of a secondary, reinforcing principle can further boost engagement, contrary to studies that seem to suggest the opposite. All in all, this heuristic selection approach remains a trial-and-error proposition at best. Once the positive average effect of persuasion over a group of users has been established, we proceed to design for individual level response (personalization) by implementing a deep learning algorithm in combination with domain knowledge, a notoriously difficult problem.

5.1.2 Maximizing User Engagement through Generalized Reinforcement Learning

To personalize the targeting of persuasion principles to individual users we implemented a novel model-free and off-policy deep reinforcement learning (DRL) agent which can be trained with a limited training dataset. We framed the task of targeting the right persuasion principle to the right user as a sequential decision making problem and applied Q-learning, one of the most traditional algorithms, to learn the optimal action-selection policy. We identified that an efficient DRL is required to have a lowered discount factor. For the algorithm to converge towards the optimal solution, the agent is presented with a low-dimensional state space representation built by compressing the limited dataset using an exponential moving average (EMA) aggregation scheme. Given that the agent can only interact with the environment in a limited way, and that the data is not truly representative of the actual state space, it could potentially suffer from high estimation variance. To reduce the variance and accommodate data sparsity, the time horizon has been limited

in such a way as to lower the Q approximation values. The EMA representation enables the regularized DQN agent to effectively capture and analyze information of state and action over time. We demonstrate through both a simulation and a real-life email marketing campaign that our methodology delivers higher engagement compared with a human expert agent using a heuristic selection approach and a control agent with randomly selected principles. The social psychology literature features numerous strategies with most of them fitting into one of six broad categories known as the universal principles of persuasion. In our study the principles of consensus (social proof), scarcity and authority result in higher engagement when embedded in the subject line of email messages.

Even though users are not sorted or characterized explicitly in the state space, a universal policy proved sufficient in personalizing the different principles to include in the next message to different users. We collected the click stream data from a historical campaign with 150 users and randomly assigned the users to RL, Human and Control groups, with 50 members in each group. To initialize the system, all users received the same message with the same persuasion principle for the first three days of the campaign and their click responses were recorded. On the fourth day, DQN begins to recommend which principles should be sent to each of the users within the RL group based on the state space of the individual users. At the same time, an expert human agent with domain knowledge makes a recommendation on the principle to send to users in the Human Expert group. Finally, users in the Control group receive a randomly selected principle.

Even though the training data was limited, prediction performance of the regularized DQN is appealing with engagement signals in the RL group approaching 300% when compared with the Control and Human Expert groups. The agent was able to train and perform well based on a few observed users' click responses. This is all the more remarkable given the shorter period of two

consecutive campaigns each lasting just 7 days. This setting mirrors real world email marketing campaigns with typically short duration of between 7 and 30 days.

Our study provides some insight into the importance of hyper-parameter tuning in the regularized RL algorithm. Specifically, we observed that the discount factor, batch size, learning rate and buffer size have the most impact on algorithm performance. It is worth noting that rather than focus this study on the best RL agent for this recommendation task, our goal was to show the power of a hybrid approach – combining domain knowledge with a data-driven methodology – to achieve promising results in terms of increased engagement over a short campaign period, utilizing small datasets. However, we note some limitations with the size of our validation set and the setting.

5.2 Future Research

5.2.1 Online Learning from Limited Samples

Despite a number of research efforts in user engagement made possible by recommender systems, studies based on RL-based recommenders are limited. Such studies tend to originate from the laboratories of big technology firms, such as Google and Facebook, at the forefront of operationalizing RL methods for ever-increasing user engagement. Given their massive data footprint, the focus and interest of their interest may not always include data and sample efficiency but that is precisely what is required for a wider adoption of RL methodologies both in the research and practice of user engagement in real world settings. There are some related works that deal with RL on real systems and at the same time focus on sample efficiency but, these studies are rarely done in the context of user engagement. More research focused on sample efficiency in the context of user engagement is highly demanded by a growing number of online businesses.

Furthermore, most real-world systems do not have separate training and evaluation environments, unlike much of the research performed in Deep RL, which rely on off-policy evaluation. To overcome the challenges of off-policy evaluation as the difference between the policies and the resulting state distribution grows, our understanding, and application, of online learning during live campaigns will take on added significance in the user engagement community.

5.2.2 Persuasive Message Composition with NLP Models

Persuasive message composition is an integral topic in online messaging, whether that be in the form of email, text, or video. How to compose such messages embedded with persuasion principles in the subject line and body copy is crucial to getting the target users to engage on a deeper level such as making purchases, referrals and return visits. With advanced transformers like GPT-2, BERT and T5 now commercially available [181], future research could investigate advanced personalization of messages and content by mimicking the customer voice, tone, and vocabulary to further boost engagement. Researchers could explore full message composition or combine natural language processing with the RL agent with the goal of reducing the time and effort it takes to create more personalized and engaging content.

5.2.3 Deploying More Sophisticated Algorithms

Our study used DQN, one of the most traditional learning algorithms, that combines Q-Learning with deep neural networks for complex, high-dimensional environments. Since our focus was not on building the "best" RL agent for this recommendation task but rather to show the power and promise of a hybrid approach in our specific data-efficient use case, we may have ended up with a slight compromise on optimality. A possible subject of future research can be the exploration of other RL algorithms, specifically actor critics such as Proximal Policy Optimization

(PPO) and Asynchronous Advantage (A3C), and comparison of results with the regularized DQN algorithm in our study. Additionally, a future research direction could include the application of our method to other communication channels such as text, chat, video as well as other use cases such as online advertising, dating and news recommendation.

REFERENCES

[1] Abhinandan S, Datar, M, Garg, A and Rajaram, S. (2007) Google news personalization: scalable online collaborative filtering. In Proceedings of the 16th international conference on World Wide Web. ACM, 271–280.

[2] Aggarwal. C., (2016) Recommender systems: The textbook. Springer, 2016.

[3] Aharony, N, Pan, W, Ip, C, Khayal, I and Pentland, A. (2011) Social fMRI: Investigating and shaping social mechanisms in the real world. Pervasive and Mobile Computing, 7(6):643–659.

[4] Amatriain X and Basilico, J. (2012) Netflix recommendations: Beyond the 5 stars. https://medium.com/net*f*ixtechblog/net*f*ix-recommendations-beyond-the-5-stars-part-1-55838468f429.

[5] Amatriain & Basilico (2012) Netflix recommendations: Beyond the 5 stars. https://medium.com/netflixtechblog/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429.

[6] Ambady, N., Bernieri, F., Richeson, J. (2000) Toward a histology of social behavior: Judgmental accuracy from thin slices of the behavioral stream, Advances in Experimental Social Psychology, Academic Press, Volume 32, 2000, Pages 201-271.

[7] Amit, R., Meir R., and Ciosek, K. (2020). "Discount Factor as a Regularizer in Reinforcement Learning." arXiv preprint arXiv:2007.02040.

[8] Ballon, P. and Schuurman, D. (2015), "Living labs: concepts, tools and cases", info, Vol. 17 No. 4. https://doi.org/10.1108/info-04-2015-0024.

[9] Ballon, P., Pierson, J. and Delaere, S. (2005), "Test and experimentation platforms for broadband innovation: examining European practice", available at: http://ssrn.com/abstract1331557

[10] Bandura, A., (1991) Social cognitive theory of self-regulation. Organ. Behav. Hum. Decis. Process.50(2),248–287.

[11] Barkhuus, L. and Rode, J.A. (2007), "From mice to men-24 years of evaluation in CHI", Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY.

[12] Barry, B., & Shapiro, D. L. (1992). Influence tactics in combination: The interactive effects of soft versus hard tactics and rational exchange. Journal of Applied Social Psychology, 22, 1429–1441.

[13] Baxter, J., and Bartlett, P. (2001) Infinite-horizon policy-gradient estimation. Journal of Artificial Intelligence Research, 15:319–350.

[14] Bergvall-Kåreborn, B., et al. (2009) A Milieu for Innovation : Defining Living Labs. Paper presented at the 2nd ISPIM Innovation Symposium, New York.

[15] Bhatnagar, S., Sutton, R., Ghavamzadeh, M. and Lee, M. (2009) "Natural actor-critic algorithms," Automatica, vol. 45, no. 11, pp. 2471–2482.

[16] Bijmolt T, Leeflang P, Block F, et al. (2010) Analytics for Customer Engagement. Journal of Service Research. 2010;13(3):341-356. doi:10.1177/1094670510375603

[17] Blass, T. (1999). The Milgram paradigm after 35 years: Some things we now know about obedience to authority. Journal of Applied Social Psychology, 29, 955-978.

[18] Brehm, J.W. (1966). A theory of psychological reactance. New York: Academic Press.

[19] Brehm, S. S., & Brehm, J. W. (1981). Psychological reactance: A theory of freedom and control. San Diego, CA: Academic Press.

[20] Brock, Timothy C. (1968), "Implications of Commodity Theory for Value Change," in Psychological Foundations of Attitudes, eds. Anthony G. Greenwald, Timothy C. Brock, and Thomas M. Ostrom, New York: Academic Press, Inc.

[22] Brown, B., Reeves, S. and Sherwood, S. (2011), “Into the wild: challenges and opportunities for field trial methods”, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM Press, New York, NY, pp. 1657-1666.

[23] Brown, B., Reeves, S. and Sherwood, S. (2011), “Into the wild: challenges and opportunities for field trial methods”, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM Press, New York, NY, pp. 1657-1666.

[24] Buckman, J., Hafner, D., Tucker, G., Brevdo, E., and Lee, H. (2018) Sample-efficient reinforcement learning with stochastic ensemble value expansion. CoRR, abs/1807.01675, 2018.

[25] Buşoniu L., Babuška R., De Schutter B. (2010) Multi-agent Reinforcement Learning: An Overview. In: Srinivasan D., Jain L.C. (eds) Innovations in Multi-Agent Systems and Applications - 1. Studies in Computational Intelligence, vol 310. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-14435-6_7

[26] Cai, H, Ren, K, Zhang, W., Malialis, K., Wang, J., Yu, Y., and Guo, D. (2017). Real-Time Bidding by Reinforcement Learning in Display Advertising. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM '17)

[27] Cesa-Bianchi, N., Gentile C., and Zappella, G. (2013). A gang of bandits. In Advances in Neural Information Processing Systems. 737–745.

[28] Chaiken, S., Liberman,A., Eagly, A.H., (1989). Heuristic and Systematic Information Processing Within and Beyond the Persuasion Context. Unintended Thought 212.

[29] Cheng, H. et. al., (2016) Wide & deep learning for recommender systems. In Proc. 1st Workshop on Deep Learning for Recommender Systems, pages 7–10.

[30] Chen, P., Chou, Y., and Kaufman, R. (2009) Community-based recommender systems: Analyzing business models from a systems operator's perspective. In Proceedings of the 42nd Hawaii International Conference on System Sciences, HICSS '09, pages 1–10, 2009.

[31] Chen, M., Beutel, A., Covington, P., Jain, S., Belletti, F. and Chi. E. (2019). Top-K Off-Policy Correction for a REINFORCE Recommender System. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19)

[32] Chen, M. et al. (2019) Top-K Off-Policy Correction for a REINFORCE Recommender System. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19). Association for Computing Machinery, New York, NY, USA, 456–464.

[33] Chua, K., Calandra, R., et al. (2018) Deep reinforcement learning in a handful of trials using probabilistic dynamics models. In Advances in Neural Information Processing Systems, pp. 4754–4765, 2018.

[34] Cialdini, R., (2001) Influence, Science and Practice. Allyn & Bacon, Boston.

[35] Cialdini, R., (2004) The science of persuasion. Sci. Am. Mind 284,76–84.

[36] Cialdini, R., (2016) Pre-suasion: a Revolutionary Way to Influence and Persuade. Simon & Schuster, New York.

[37] Cialdini, R., (2021). Influence, the Psychology of Persuasion. New and Expanded. Harper Business.

[38] Cobbe, K., Klimov, O., Hesse, C., Kim, T., & Schulman, J. (2019). Quantifying generalization in reinforcement learning. In International Conference on Machine Learning (pp. 1282-1289). PMLR.

[39] Coorevits, L., Georges, A., & Schuurman, D. (2018) A Framework for Field Testing in Living Lab Innovation Projects. Technology Innovation Management Review, 8(12): 40-50. http://doi.org/10.22215/timreview/1204.

[40] Coorevits, L. et al., (2018) A Framework for Field Testing in Living Lab Innovation Projects. Technology Innovation Management Review, 8(12): 40-50. http://doi.org/10.22215/timreview/1204

[41] Covington, P., Adams, J., and Sargin, E. (2016) Deep neural networks for YouTube recommendations. In Proceedings of the 10th ACM conference on recommender systems, pp. 191–198. ACM, 2016.

[42] Curhan, J. R., & Pentland, A. (2007). Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes. Journal of Applied Psychology, 92(3), 802–811.

[43] Morales, G., Gionis, A., & Lucchese, C. (2012). From chatter to headlines: harnessing the real-time web for personalized news recommendation. In Proceedings of the fifth ACM international conference on Web search and data mining (pp. 153-162).

[44] Davidson,J., Liebald, B., Liu, J., Nandy, P., Vleet, T., Gargi, U, Gupta, S., et al. (2010) The YouTube Video Recommendation System. In Proceedings of the Fourth ACM Conference on Recommender Systems, RecSys '10, pages 293–296, 2010.

[45] Deisenroth, M. P., and Rasmussen, C. E. (2011). PILCO: A Model-Based and Data-Efficient Approach to Policy Search. In Proceedings of the International Conference on Machine Learning.

[46] Dekker, R, Contreras, F., Meijer, A. (2019) The living lab as a methodology for public administration research: A systematic literature review of its applications in the social sciences. International Journal of Public Administration 43(14): 1207–17.

[47] Deutsch, M., & Gerard, H. (1955). A study of normative and informational social influences upon individual judgment. The Journal of Abnormal and Social Psychology, 51(3), 629–636. https://doi.org/10.1037/h0046408

[48] Doorn J, Lemon, K, Mittal V, et al. (2010) Customer Engagement Behavior: Theoretical Foundations and Research Directions. Journal of Service Research. 2010;13(3):253-266. doi:10.1177/1094670510375599

[49] Dulac-Arnold, G., Evans, R., van Hasselt, H., Sunehag, P., Lillicrap, T., Hunt, J., Mann, T., Weber, T., Degris, T., and Coppin, B. (2015) Deep reinforcement learning in large discrete action spaces. arXiv preprint arXiv:1512.07679, 2015.

[50] Dulac-Arnold, G., Mankowitz, D., and Hester, T. (2019) Challenges of real-world reinforcement learning. arXiv preprint arXiv:1904.12901, 2019.

[51] Eagle, N., Pentland, A. (2006) Reality mining: sensing complex social systems, Personal and Ubiquitous Computing (10) (2006) 255–268.

[52] Eagle, N., Pentland, A., Lazer, D. (2009) Inferring friendship network structure by using mobile phone data, Proceedings of the National Academy of Sciences 106 (36) (2009).

[53] Eisend, M. (2008) Explaining the impact of scarcity appeals in advertising: the mediating role of perceptions of susceptibility. J. Advert. 37 (3), 33–40.

[54] Emarsys.com (2016) 'Emarsys Survey Finds SMBs Quickly Adapting to the Omnichannel Paradigm to Compete | Available from: https://emarsys.com/press-release/emarsys-survey-finds-smbs-quickly-adapting-omnichannel-paradigm-compete/ (accessed 14 October 2019).

[55] Eriksson, M., Niitamo, V and Kulkki, S. (2005). State-of-the-art in Utilizing Living Labs Approach to User-centric ICT innovation - a European approach:

[56] Falbe, C. M., & Yukl, G. (1992). Consequences for managers of using single influence tactics and combinations of tactics. Academy of Management Journal, 35, 638-653.

[57] Festinger, L., (1957) A Theory of Cognitive Dissonance. Stanford University Press, Stanford.

[58] Finn, C., Abbeel, P., and Levine, S. (2017) Model-agnostic meta learning for fast adaptation of deep networks. In Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp. 1126–1135. JMLR. org, 2017.

[59] Fishbein, M., Ajzen, I., (2011) Predicting and Changing Behavior: The Reasoned Action Approach. Taylor & Francis, London.

[60] Freedman, J. L., & Fraser, S. C. (1966). Compliance without pressure: The foot-in-the-door technique. Journal of Personality and Social Psychology, 4(2), 195–202. https://doi.org/10.1037/h0023552

[61] Fogg, B. (2003) Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufmann, San Francisco, CA.

[62] Fogg BJ. (2020) Tiny Habits: The Small Changes That Change Everything. Houghton Miin Harcourt, Boston, MA.

[63] Garner R., (2005) Post-It® note persuasion: a sticky influence, Journal of Consumer Psychology, 15(3): 230–7.

[64] Goh, K. Y., Heng, C. S., & Lin, Z. (2013). Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. Information Systems Research, 24 (1), 88–107.

[65] Goldberg, D. Nichols, D., Oki, B. M., and Terry, D. (1992) Using collaborative filtering to weave an information tapestry. Communications of the ACM, 1992 - dl.acm.org

[66] Goldstein, N.J., Cialdini, R.B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. Journal of Consumer Research, 35(3), 472–482.

[67] Gomez-Uribe, C. and N. Hunt. (2015). The Netflix recommender system: Algorithms, business value, and innovation. Transactions on Management Information Systems, 6(4):13:1–13:19.

[68] Greenberg M.S. (1980) A Theory of Indebtedness. In: Gergen K.J., Greenberg M.S., Willis R.H. (eds) Social Exchange. Springer, Boston, MA. https://doi.org/10.1007/978-1-4613-3087-5_1.

[69] Guanjie Z, Fuzheng Z, Zihan Z, Yang X, et al. (2018). DRN: A Deep Reinforcement Learning Framework for News Recommendation. In www. 167–176.

[70] Guo, H, Tang, R, Ye, Y, Li, Z. and He, X. (2017) DeepFM: a factorization machine based neural network for CTR prediction. arXiv preprint arXiv:1703.04247.

[71] van Hasselt, H., Guez, A., & Silver, D. (2016). Deep Reinforcement Learning with Double Q-Learning. Proceedings of the AAAI Conference on Artificial Intelligence, 30(1). Retrieved from https://ojs.aaai.org/index.php/AAAI/article/view/10295

[72] He, J., Chen, J., He, X., Gao, J., Li, L., Deng, L., and Ostendorf, M. (2015) Deep reinforcement learning with a natural language action space. arXiv preprint arXiv:1511.04636, 2015.

[73] Heider, F. (1958). The psychology of interpersonal relations. New York: Wiley

[74] Hester, T. and Stone, P. (2013) TEXPLORE: Realtime sample-efficient reinforcement learning for robots. Machine Learning, 90(3), 2013.

[75] Hornstein, H. A., Fisch, E., & Holmes, M. (1968). Influence of a model's feeling about his behavior and his relevance as a comparison other on observers' helping behavior. Journal of Personality and Social Psychology, 10(3), 222–226. https://doi.org/10.1037/h0026568

[76] Hubspot.com (2020) 'Not Another State of Marketing Report' p. 42 | Available from: https://www.hubspot.com/state-of-marketing (accessed 4 February 2021)

[77] IJntema, W., Goossen, F., Frasincar, F. and Hogenboom, F (2010). Ontology-based news recommendation. In Proceedings of the 2010 EDBT/ICDT Workshops. ACM, 16.

[78] Inman, J. J., Peter, A. C., & Raghubir, R. P. (1997). Framing the deal: The role of restrictions in accentuating deal value. Journal of Consumer Research, 24, 68–79, (June).

[79] Irpan, A. (2018) Deep reinforcement learning doesn't work yet. https://www.alexirpan.com/2018/02/14/rl-hard.html

[80] Jaakkola E. & Alexander M. (2014) The Role of Customer Engagement Behavior in Value Co-Creation: A Service System Perspective. Journal of Service Research. 2014;17(3):247-261. doi:10.1177/1094670514529187

[81] Jannach, D., & Jugovac, M. (2019). Measuring the business value of recommender systems. ACM Transactions on Management Information Systems (TMIS), 10(4), 1-23.

[82] Jordan Larson, (2014) "The invisible manipulative power of persuasive technology," Pacific Standard, May 14, 2014. https://psmag.com/environment/captology-fogg-invisible-manipulative-power-persuasive-technology-81301.

[83] Kaasinen, E., Koskela-Huotari, K., et al. (2013), “Three approaches to co-creating services with users,” Advances in the Human Side of Service Engineering, Vol. 286.

[84] Kahneman, D and Tversky, A. (1979). "Prospect Theory: An Analysis of Decision Under Risk," Econometrica 47, 263-291.

[85] Kakade, S. M. (2002) A natural policy gradient. In Advances in neural information processing systems, pp. 1531–1538, 2002.

[86] Kaptein, M.C., de Ruyter, B., Markopoulos , P., Aarts, E., (2012) Tailored persuasive text messages to reduce snacking. Trans. Interact. Intell. Syst.2 (2), 10–35.

[87] Kaptein M, De Ruyter B, Markopoulos P, Aarts E. (2012) Adaptive persuasive systems: a study of tailored persuasive text messages to reduce snacking. ACM Trans Interact Intell Syst 2(2).

[88] Kaptein, M.C., van Halteren, A., (2013) Adaptive persuasive messaging to increase service retention. J. Personal Ubiquitous Computing. 17 (6), 1173–1185.

[89] Kaptein, M. C., and Kruijswijk, J. (2016). Streamingbandit: Developing adaptive persuasive systems. arXiv preprint arXiv:1602.06700.

[90] Kaptein, M., Markopoulos, P., De Ruyter, B., and Aarts, E. (2015). Personalizing persuasive technologies: Explicit and implicit personalization using persuasion profiles.

International Journal of Human Computer Studies 77: 38–51. http://doi.org/10.1016/j.ijhcs.2015.01.004

[91] Katukuri, J, Konik, T., Mukherjee, R. and Kolay, S. (2014) Recommending similar items in large-scale online marketplaces. In IEEE International Conference on Big Data 2014, pages 868–876, 2014.

[92] Kellermann & Cole, (1994) Classifying compliance gaining messages: taxonomic disorder and strategic confusion. Commun. Theory 4 (1),3–60.

[93] Kjeldskov, J. and Skov, M.B. (2014), "Was it worth the hassle? Ten years of mobile HCI research discussions on lab and field evaluations", Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services, ACM, New York, NY, pp. 43-52.

[94] Knowles, E.S., Linn, J.A., (2004) Resistance and Persuasion. Psychology Press, Abingdon.

[95] Kompan M and Bieliková, M. (2010) Content-Based News Recommendation. In EC-Web, Vol. 61. Springer, 61–72.

[96] Konda, V. R. and Tsitsiklis, J. N. (2003) "On actor–critic algorithms", SIAM Journal on Control and Optimization, 42(4):1143-1166.

[97] Korn, M. and Bødker, S. (2012), "Looking ahead: how field trials can work in iterative and exploratory design of ubicomp systems", Proceedings of the 2012 ACM Conference on Ubiquitous Computing, ACM, New York, NY, pp. 21-30.

[98] Kröse, B. (1995) Learning from delayed rewards. Journal of Robotics and Autonomous Systems, 1995, 15(4):233-235.

[99] Kumar, V., Aksoy, L., et al. (2010). Undervalued or overvalued customers: Capturing total customer engagement value. Journal of Service Research, 13(3), 297-310.

[100] Lagoudakis, M. G. and Parr, R. (2003). Least-squares policy iteration. JMLR, 4:1107–1149.

[101] Lazer, D., Pentland, A. et al., (2009) Computational social science, Science 323 (5915) (2009) 721–723. http://www.sciencemag.org/content/323/5915/721.

[102] Lee, D. and Hosanagar, K. (2014) Impact of recommender systems on sales volume and diversity. In Proceedings of the 2014 International Conference on Information Systems, ICIS ’14, 2014.

[103] Li, L., Chu, W., Langford, J. and Schapire, R. (2010) A contextual bandit approach to personalized news article recommendation. In Proceedings of the 19th international conference on World wide web. ACM, 661–670.

[104] Li, L., Wang, D., Li, T., Knox, D., and Padmanabhan, B. (2011) SCENE: a scalable two-stage personalized news recommendation system. In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, 125–134.

[105] Lillicrap, T. P., et al. (2015). Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971.

[106] Lindstrom T., Middlecamp, C. (2017) Campus as a living laboratory for sustainability: the chemistry connection. J Chem Educ 94:1036–1042. ttps://doi.org/10.1021/acs.jchemed.6b00624

[107] Litmus.com (2020) '2020_State_of_Email_Engagement | Available from: https://www.litmus.com/blog/2020-state-of-email-report-the-beginning-of-a-new-email-decade/ (accessed 12 February 2020)

[108] Liu, J., Dolan, P., Pedersen, E (2010) Personalized news recommendation based on click behavior. In Proceedings of the 15th international conference on Intelligent user interfaces. ACM, 31–40.

[109] Zou, L., Xia, L., Ding, Z., Song, J., Liu, and Yin, D (2019) Reinforcement Learning to Optimize Long-term User Engagement in Recommender Systems. In KDD. 2810–2818

[110] Long, J.D., Stevens, K.R., (2004) Using Technology to Promote Self-Efficacy for Healthy Eating in Adolescents. Technical Report 2, Lubbock Christian University, 5601 W.19th Street, Lubbock, TX 79407, USA[111]

[111] Lu, Z. and Yang, Q. (2016) Partially Observable Markov Decision Process for Recommender Systems. arXiv preprint arXiv:1608.07793 (2016).112]

[112] Mahmood T and Ricci F (2009) Improving recommender systems with adaptive conversational strategies. In HT'09. ACM, 73–82.

[113] Marcus, G. (2018). Deep learning: A critical appraisal. arXiv preprint arXiv:1801.00631.

[114] Margalit, L. (2019), "The Psychology of Customer Experience", Einav, G. (Ed.) Digitized, Emerald Publishing Limited, Bingley, pp. 87-98. https://doi.org/10.1108/978-1-78973-619-920191005

[115] Marlin B and Zemel, R. (2004) The multiple multiplicative factor model for collaborative filtering. In Proceedings of the twenty-first international conference on Machine learning. ACM, 73.

[116] Maslow, A., Herzeberg, A., (1954) Hierarchy of needs. In: Maslow, A.H. (Ed.), Motivation and Personality. Harper, New York.

[117] Milgram, S. (1974). Obedience to authority. New York: Harper & Row.

[118] Mims, C. (2019) "The Hot New Channel for Reaching Real People: Email; Frustrated by Social Media, Businesses and Others Looking for an Audience Turn to an Old Standby." Wall Street Journal (Online), Jan 19 2019, ProQuest. Web. 6 Apr. 2021.

[119] MIT Technology Review (2021) 'How Facebook got addicted to spreading misinformation | MIT Technology Review', Available from: https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/ (accessed 21 March 2021).

[120] Mnih, V., Kavukcuoglu, et al. (2013) Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602, 2013.

[121] Mnih, V., Kavukcuoglu, D. Silver, A. et al. (2015). "Human-level control through deep reinforcement learning." nature 518(7540): 529-533.

[122] Mnih, V., Kavukcuoglu, K., Silver, D. et al. (2015) Human-level control through deep reinforcement learning. Nature 518, 529–533 (2015). https://doi.org/10.1038/nature14236

[123] Moe, W. M., & Trusov, M. (2011). The value of social dynamics in online product ratings forums. Journal of Marketing Research, 48(3), 444–456.

[124] Moore, A. and Atkeson, C. G. (1993). Prioritized sweeping: Reinforcement learning with less data and less time. Machine Learning, 13:103–130.

[125] Naumov, M. et al., (2019) "Deep learning recommendation model for personalization and recommendation systems," CoRR, vol. abs/1906.00091, 2019. [Online]. Available: http://arxiv.org/abs/1906. 00091 [39]

[126] Osband, I., Blundell, C., et al. (2016) Deep exploration via bootstrapped DQN. Advances in Neural Information Processing Systems 29, pp. 4026–4034. Curran Associates, Inc., 2016.

[127] Phelan, O., McCarthy, K., Bennett, M., Smyth, B (2011). Terms of a feather: Content-based news recommendation and discovery using twitter. Advances in Information Retrieval (2011), 448–459.

[128] Paredes P, Gilad-Bachrach R, et al. (2014) Pop Therapy: Coping with Stress Through Pop-culture. In: Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare.

[129] Park, S., Shim, et al. (2016) Multimodal Analysis and Prediction of Persuasiveness in Online Social Multimedia. ACM Trans. Interact. Intell. Syst. 6, 3, Article 25 (October 2016)

[130] Peters J., Vijayakumar S., Schaal S. (2005) Natural Actor-Critic. In: Gama J., Camacho R., Brazdil P.B., Jorge A.M., Torgo L. (eds) Machine Learning: ECML 2005. ECML 2005. Lecture Notes in Computer Science, vol 3720. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11564096_29

[131] Peters, J., Schaal, S. (2008) Natural Actor-Critic, Neurocomputing, Volume 71, Issues 7–9, 2008, Pages 1180-1190, https://doi.org/10.1016/j.neucom.2007.11.026.

[132] Petty & Cacioppo, (1986) The elaboration likelihood model of persuasion. Adv. Exp. Soc.Psychol. 19(1), 123–205.

[133] Prochaska, J.O., Velicer, W.F., (1997) The transtheoretical model of health behavior change. Am.J. Health Promot.12 (1), 38–48.

[134] Radicati Group (2021) 'Email Statistics Report, 2021-2025 | Available from: https://www.radicati.com/?p=17245 (accessed 27 March 2021)

[135] Riedmiller, M. (2005) Neural Fitted Q Iteration – First Experiences with a Data Efficient Neural Reinforcement Learning Method. In: Gama J., Camacho R., Brazdil P.B., Jorge A.M., Torgo L. (eds) Machine Learning: ECML 2005.

[136] Redlich, K.C. (2016) 47 Proven Headlines That Work Like Crazy. https://medium.com/@carlosredlich/on-the-average-five-times-as-many-people-read-the-headlines-as-read-the-body-copy-48501d7a3439

[137] Rendle, S. (2010) Factorization machines. In Proc. 2010 IEEE International Conference on Data Mining, pages 995–1000.

[138] Resnick, P. and Varian, H. (1997) Recommender systems, Comm. ACM, 40, pp. 56–58.

[139] ReturnPath (2017) Email Marketing Performance in 2017 | Available from: https://returnpath.com/newsroom/marketing-executives-say-email-performance-rise-personalization-effective-email-marketing-tactic/

[140] Rhoads, K. (2007) How Many Influence, Persuasion, Compliance Tactics & Strategies Are There? ⟨http://www.workingpsychology.com/numbertactics.html⟩.

[141] Ricci, F., Rokach, L., Shapira, B. and Kantor, P. (2011) Recommender systems handbook. Springer, New York.

[142] Rodriguez, M., Posse, C., and Zhang, E. (2012) Multiple objective optimization in recommender systems. In Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys ’12, pages 11–18, 2012.

[143] Rodriguez, M., Posse, C., and Zhang, E. (2012) Multiple objective optimization in recommender systems. In Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys ’12, pages 11–18, 2012.

[144] Romero, Herrera (2017) The Emergence of Living Lab Methods. In: Keyson D., Guerra-Santin O., Lockton D. (eds) Living Labs. Springer, Cham. https://doi.org/10.1007/978-3-319-33527-8_2

[145] Rummery, G., & Niranjan, M. (1994). On-line Q-learning using connectionist sytems (Technical Report CUED/FINFENG-TR 166). Cambridge University, UK.

[146] Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist, 55, 68–78.

[147] Sakai, R., van Peteghem, S., vande Sande,L., Banach, P., Kaptein,M.C., (2011) Personalized persuasion in ambient intelligence: the APStairs system. In: Proceedings of Ambient Intelligence (AmI) 2011, Amsterdam.

[148] Sánchez-Corcuera R., Casado-Mansilla D., Borges C.E., López-de Ipiña D. (2020) Persuasion-based recommender system ensambling matrix factorisation and active learning models Pers. Ubiquitous Comput. (2020), pp. 1-11

[149] Sánchez-Corcuera, R. et al. (2020) Persuasion-based recommender system ensambling matrix factorisation and active learning models. Pers Ubiquit Comput (2020). https://doi.org/10.1007/s00779-020-01382-7

[150] Sánchez-Corcuera R, Nuñez-Marcos A, Sesma-Solance J, et al. (2019) Smart cities survey: Technologies, application domains and challenges for the cities of the future. International Journal of Distributed Sensor Networks. June 2019. doi:10.1177/1550147719853984

[151] Schelling, T. (1978). Egonomics, or the Art of Self-Management. The American Economic Review, 68(2), 290-294. Retrieved April 7, 2021, from http://www.jstor.org/stable/1816707

[152] Schrage, M. (2021) "The Transformational Power of Recommendation." MIT Sloan Management Review 62.2 (2021): 17-21. ProQuest. Web. 6 Apr. 2021.

[153] Schrage, M. (2020) Recommendation Engines. MIT Press.

[154] Schulman,J., Moritz,P., et al. (2015) “High-dimensional continuous control using generalized advantage estimation”. In: arXiv preprint arXiv:1506.02438 (2015).

[155] Schumacher and Feurstein, (2007) "Living Labs - the user as co-creator," 2007 IEEE International Technology Management Conference (ICE), Sophia Antipolis, France, 2007, pp. 1-6.

[156] Schuurman, D. et al. (2016) The Impact of Living Lab Methodology on Open Innovation Contributions and Outcomes. Technology Innovation Management Review, 6(1): 7-16. http://doi.org/10.22215/timreview/956

[157] Sein, M. K., Henfridsson, O., Rossi, M., & Lindgren, R. (2011) Action Design Research. MIS Quarterly, 35(1): 37–56.

[158] Seta, J.J. and Seta, C.E. (1992), “Personal equity-comparison theory: an analysis of value and the generation of compensatory and noncompensatory expectancies,” Basic and Applied Social Psychology, Vol. 13, March, pp. 47-66.

[159] Silver, D. et al. (2016) Mastering the game of Go with deep neural networks and tree search. Nature 529, 484–489 (2016).

[160] Simon, H. (1972). Theories of Bounded Rationality. In Decision and Organization, edited by C. B. Radner and Roy Radner, 161–76. Amsterdam: North-Holland.

[161] Skinner, B.F., (1976) About Behaviorism. Vintage Books, New York.

[162] Smith & Linden, (2017) "Two Decades of Recommender Systems at Amazon.com," in IEEE Internet Computing, vol. 21, no. 3, pp. 12-18, May-June 2017, doi: 10.1109/MIC.2017.72.

[163] Smyth, B., CoŠer, P., and Oman, S. (2007) Enabling intelligent content discovery on the mobile internet. In Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence, AAAI ’07, pages 1744–1751.

[164] Søndergaard, M., Karnøe, M., Nelson, M., Fogg, BJ (2013) “Persuasive Business Model,” Journal of Multi Business Model Innovation and Technology, vol. 1, pp. 71–100, 2013.

[165] Statista.com (2021) 'Change in user engagement with selected social media platforms in the United States from 2018 to August 2020 | Available from: https://www.technologyreview.com/2021/03/11/1020600/facebook-responsible-ai-misinformation/ (accessed 30 January 2021)

[166] Ståhlbröst, A., Bertoni, M., et al. (2013), “Social media for user innovation in Living Labs: a framework to support user recruitment and commitment”, XXIV ISPIM Conference – Innovating in Global Markets: Challenges for Sustainable Growth, Helsinki.

[167] Sutton, R. (1991). Dyna, an integrated architecture for learning, planning, and reacting. SIGART Bull. 2, 4 (Aug. 1991), 160–163. DOI: https://doi.org/10.1145/122344.122377

[168] Szpektor, I., Maarek, Y., and Pelleg, D. (2013) When relevance is not enough: Promoting diversity and freshness in personalized question recommendation. In Proceedings of the 22nd International Conference on World Wide Web, WWW ’13, pages 1249–1260, 2013.

[169] Tang, L. et al. (2015). Personalized recommendation via parameter-free contextual bandits. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 323–332.

[171] Tang, L. et al. (2014). Ensemble contextual bandits for personalized recommendation. In Proceedings of the 8th ACM Conference on Recommender Systems. ACM, 73–80.

[172] Thaler, R. (2017) “Behavioral Economics,” Journal of Political Economy, Vol. 125 (December), pp. 1799–1805.

[173] Varaiya, P et al. (1983) "Extension of the multi-armed bandit problem," The 22nd IEEE Conference on Decision and Control, San Antonio, TX, USA, 1983, pp. 1179-1180, doi: 10.1109/CDC.1983.269708.

[174] Verhoef, P., Reinartz, W., Krafft, M. (2010) Customer Engagement as a New Perspective in Customer Management. Journal of Service Research. 2010;13(3):247-252. doi:10.1177/1094670510375461

[175] Vroom, V. (1964). Work and motivation. New York, NY: Wiley & Sons.

[176] Wang, H., Chen, X., et al. (2017). Integrating Reinforcement Learning with Multi-Agent Techniques for Adaptive Service Composition. ACM Transactions on Autonomous and Adaptive Systems. 12. 1-42. 10.1145/3058592.

[177] Wang, X et al. (2014). Exploration in interactive personalized music recommendation: a reinforcement learning approach.

[178] Watkins, C. & Dayan, P. (1992) Q-learning. Mach. Learn. 8, 279–292

[179] West, S. G. (1975), “Increasing the Attractiveness of College Cafeteria Food: A Reactance Theory Perspective,” Journal of Applied Psychology, 60 (5), 656–658.

[180] Williams, R. (1992) Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning, 8(3-4):229–256.

[181] Wolf, T., Debut, L., Sanh, V. et al. (2019). Hugging Face's Transformers: State-of-the-art natural language processing. arXiv preprint arXiv:1910.03771.

[182] Yin, R. K. (2009) Case Study Research: Design and Methods. Thousand Oaks, CA: Sage.

[183] Yoon, G., Li, C., Ji, Y., North, M., Hong, C., & Liu, J. (2018). Attracting comments: Digital engagement metrics on Facebook and financial performance. Journal of Advertising, 47(1), 24-37.

[184] Zahavy, T., Haroush,M., Merlis, N., Mankowitz, D. J., and Mannor, S. (2018) Learn what not to learn: Action elimination with deep reinforcement learning. In Advances in Neural Information Processing Systems, pp. 3562–3573, 2018.

[185] Zeng,C. et al. (2016). Online Context-Aware Recommendation with Time Varying Multi-Armed Bandit. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2025–2034.

[187] Zhao, C., Sigaud, O., et al. (2019). Investigating generalisation in continuous deep reinforcement learning. arXiv preprint arXiv:1902.07015.

[188] Zhao, X. et al. (2019). "Deep reinforcement learning for search, recommendation, and online advertising: a survey" SIGWEB Newsl., Spring, Article 4 (Spring 2019), 15 pages.

[189] Zhao, X. et al. (2018). Deep Reinforcement Learning for Search, Recommendation, and Online Advertising: A Survey. arXiv preprint arXiv:1812.07127.

[190] Zheng, G. et al. (2018) DRN: A Deep Reinforcement Learning Framework for News Recommendation. In Proceedings of the 2018 World Wide Web Conference (WWW '18)

[191] Zhu, F., and Zhang, X. (2010) Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Journal of Marketing, 74 (March 2010), 133-148.

[192] Redlich, K.C. (2016) 47 Proven Headlines That Work Like Crazy. https://medium.com/@carlosredlich/on-the-average-five-times-as-many-people-read-the-headlines-as-read-the-body-copy-48501d7a3439

ABSTRACT

MAXIMIZING USER ENGAGEMENT IN SHORT MARKETING CAMPAIGNS WITHIN AN ONLINE LIVING LAB: A REINFORCEMENT LEARNING PERSPECTIVE

by

ANIEKAN MICHAEL INI-ABASI

August 2021

Advisor: Dr. Ratna Babu Chinnam

Major: Industrial & Systems Engineering

Degree: Doctor of Philosophy

User engagement has emerged as the engine driving online business growth. Many firms have pay incentives tied to engagement and growth metrics. These corporations are turning to recommender systems as the tool of choice in the business of maximizing engagement. LinkedIn reported a 40% higher email response with the introduction of a new recommender system. At Amazon 35% of sales originate from recommendations, while Netflix reports that '75% of what people watch is from some sort of recommendation,' with an estimated business value of $1 billion per year. While the leading companies have been quite successful at harnessing the power of recommenders to boost user engagement across the digital ecosystem, small and medium businesses (SMB) are struggling with declining engagement across many channels as competition for user attention intensifies. The SMBs often lack the technical expertise and big data infrastructure necessary to operationalize recommender systems.

The purpose of this study is to explore the methods of building a learning agent that can be used to personalize a persuasive request to maximize user engagement in a data-efficient setting. We frame the task as a sequential decision-making problem, modelled as MDP, and solved using

a generalized reinforcement learning (RL) algorithm. We leverage an approach that eliminates or at least greatly reduces the need for massive amounts of training data, thus moving away from a purely data-driven approach. By incorporating domain knowledge from the literature on persuasion into the message composition, we are able to train the RL agent in a sample efficient and operant manner.

In our methodology, the RL agent nominates a candidate from a catalog of persuasion principles to drive higher user response and engagement. To enable the effective use of RL in our specific setting, we first build a reduced state space representation by compressing the data using an exponential moving average scheme. A regularized DQN agent is deployed to learn an optimal policy, which is then applied in recommending one (or a combination) of six universal principles most likely to trigger responses from users during the next message cycle. In this study, email messaging is used as the vehicle to deliver persuasion principles to the user. At a time of declining click-through rates with marketing emails, business executives continue to show heightened interest in the email channel owing to higher-than-usual return on investment of $42 for every dollar spent when compared to other marketing channels such as social media.

Coupled with the state space transformation, our novel regularized Deep Q-learning (DQN) agent was able to train and perform well based on a few observed users' responses. First, we explored the *average* positive effect of using persuasion-based messages in a live email marketing campaign, without deploying a learning algorithm to recommend the influence principles. The selection of persuasion tactics was done heuristically, using only domain knowledge. Our results suggest that embedding certain principles of persuasion in campaign emails can significantly increase user engagement for an online business (and have a positive impact on revenues) without putting pressure on marketing or advertising budgets. During the study, the store had a customer

retention rate of 76% and sales grew by a half-million dollars from the three field trials combined. The key assumption was that users are predisposed to respond to certain persuasion principles and learning the right principles to incorporate in the message header or body copy would lead to higher response and engagement.

With the hypothesis validated, we set forth to build a DQN agent to recommend candidate actions from a catalog of persuasion principles most likely to drive higher engagement in the next messaging cycle. A simulation and a real live campaign are implemented to verify the proposed methodology. The results demonstrate the agent's superior performance compared to a human expert and a control baseline by a significant margin (~ up to 300%). As the quest for effective methods and tools to maximize user engagement intensifies, our methodology could help to boost user engagement for struggling SMBs without prohibitive increase in costs, by enabling the targeting of messages (with the right persuasion principle) to the right user.

AUTOBIOGRAPHICAL STATEMENT

Aniekan Michael Ini-Abasi received his B.Sc. degree in Computer Science from the University of Nigeria, Nsukka, Nigeria, in 1985. For two decades he worked as a software developer and later as a project lead for enterprise resource planning (ERP) initiatives in the banking industry and in municipal government. He received his MMSc degree in Management Science from the University of Waterloo, Ontario, Canada in 2010. During his master's work, he founded two startup companies focused on building systems for customer acquisition and engagement in the marketing technologies (MarTech) space. During his doctoral study in the Industrial and Systems Engineering GET program at Wayne State University, he has deepened his knowledge on product development, data-driven systems, and the application of artificial intelligence principles to maximizing user engagement on digital platforms.