Understanding Ordinary Goal Pursuit: Describing and Predicting Success in New Year's Resolutions

by

Hannah Moshontz de la Rocha

Department of Psychology and Neuroscience
Duke University

Date:_______________________
Approved:

____
Rick H. Hoyle, Advisor

____
Elizabeth J. Marsh

____
James Y. Shah

____
Gráinne M. Fitzsimons

Dissertation submitted in partial fulfillment of
the requirements for the degree of Doctor
of Philosophy in the Department of
Psychology and Neuroscience in the Graduate School
of Duke University

2020

ABSTRACT

Understanding Ordinary Goal Pursuit: Describing and Predicting Success in New Year's Resolutions

By

Hannah Moshontz de la Rocha

Department of Psychology and Neuroscience
Duke University

Date:_______________________
Approved:

____
Rick H. Hoyle, Advisor

____
Elizabeth J. Marsh

____
James Y. Shah

____
Gráinne M. Fitzsimons

An abstract of a dissertation submitted in partial
fulfillment of the requirements for the degree
of Doctor of Philosophy, Master of Science, Master of Arts in the Department of
Psychology and Neuroscience in the Graduate School of
Duke University

2020

Abstract

People's success in achieving their goals can have profound consequences for their subjective and objective well-being. Hundreds of research studies identify factors associated with success in goal pursuit, but little is known about the occurrence and influence of these factors in daily life. This dissertation aims to complement and build on extant, mostly laboratory, research by characterizing ordinary goal pursuit and identifying factors that meaningfully affect it in the context of daily life. The first chapter offers background: a review of prior research, a discussion of potential limitations on the replicability and generalizability of prior research, and an argument for more robust, naturalistic, and descriptive work. The chapters that follow present prospective observational studies focused on pursuit of New Year's resolutions and used to address eight research questions pertaining to the content and framing of goals people pursue, the outcomes of goal pursuit, and the potentially mutable factors associated with goal achievement. The second chapter presents Study 1, a descriptive study focused on understanding what goals people set as resolutions and the typical process and outcome of pursuit. The third chapter presents Study 2, a study focused on assessing the predictive value of goal-varying factors. Goals varied greatly in their content, properties, and outcomes. Contrary to theory, many resolutions were neither successful nor unsuccessful, but instead were still being pursued or were on hold at the end of the year. Across both studies, the three most common resolution outcomes at the end of the year were achievement (estimates ranged from 20% to 40%), continued pursuit (32% to 60%) and pursuit put on hold (15% to 21%). Other outcomes (e.g., deliberate disengagement) were rare (<1% to 3%). Motivation and habit formation were associated with subjective success consistently, over and above trait self-control, but no other goal-varying properties showed robust associations with goal outcomes. Predictive models suggest that relatively little variance in goal outcomes can be meaningfully predicted by goal-varying properties, and that linear regression models are particularly bad at predicting goal outcomes. This dissertation demonstrates the value of naturalistic, descriptive, and prediction-focused work for advancing understanding of self-regulation.

Contents

List of Tables

List of Figures

Acknowledgements

Thank you to my committee members who have guided me as I've developed this project and its predecessor, my MAP. Thank you to my graduate advisor and mentor, Rick Hoyle, who welcomed me into his lab, always greeted my interests with encouragement, and taught me the joys and challenges of modeling (and trying to understand) self-regulatory processes. Thank you to my secondary mentor, Beth Marsh, who welcomed me into her lab, too, and provided expert guidance in our empirical projects and all else. Thank you to Gráinne Fitzsimons and Jim Shah, for lending their expertise and always providing constructive criticism and feedback.

Thank you to the many other mentors who have shaped my career, including Kathy Oleson, Harris Cooper, Mark Leary, Madeline Carrig, Keith Payne, and present and past members of the Hoyle Lab and the Marsh Lab.

Thank you to my friends and family, and especially to my parents, Raquelle, Dan, and Chuck.

1. Introduction

How people manage their behavior, thoughts, and emotions in service of their goals (*self-regulation*) affects whether or not they achieve their goals and, by extension, their subjective and objective wellbeing. The feeling of making progress towards goals is associated with increased subjective well-being, especially when people feel they are making progress on personally or culturally meaningful goals (Brunstein, 1993; Emmons, 1986; King et al., 1998; Oishi & Diener, 2001, 2001). Effective self-regulation not only feels good, it is also associated with more tangible beneficial outcomes. Self-regulation can have profound implications for health and health behaviors, like medication adherence (e.g., Bosworth et al., 2018). Up to 50% of deaths in the United States are preventable by effective pursuit of goals like quitting cigarettes and exercising regularly (Danaei et al., 2009; Medicine, 2015; Micha et al., 2017). People skilled in self-regulation not only have happier lives, they also have longer lives (Friedman et al., 1995; Robson et al., 2020).

This dissertation aims to advance collective knowledge about effective self-regulation in the context of ordinary goal pursuit in a sample of American adults. The two studies presented here are prospective longitudinal surveys of people's New Year's resolutions. These studies build on and complement a long and varied list of theoretical and analytic traditions but are primarily situated in social psychological research on self-regulation and its mechanisms (e.g., deliberate goal pursuit). This is reflected most obviously in the focus on properties of goals and on variable contextual factors in the goal pursuit environment that affect goal outcomes.

However, several features distinguish the empirical research in this dissertation from typical social psychological research on goal pursuit and self-regulation: its ecological validity, scope (with respect to time, number of goals, and the number of predictors considered), and use of theory-based, but theoretically agnostic, analytic approaches. These studies capture the entire lifespan of long-term goals prospectively; they begin in the first days of goal pursuit and end after the ostensible intended deadline of pursuit (either the date of intended goal completion or, for maintenance goals, pursuit stability). In addition, these studies are unique from all other prior studies of New Year's resolutions in allowing people to describe goals in their own language and in assessing the influence of individual differences separately from the influence of goal-varying properties, something that was made possible by allowing people to report multiple resolutions. Both individual differences and properties of goals have long been recognized as core mechanisms of effective self-regulation, as elaborated on in the next section.

1.1 Mechanisms of effective self-regulation

Prior research has identified two kinds of mechanisms of effective self-regulation: individual differences and properties of goals (and goal pursuits). People reliably differ in their self-regulatory tendencies; some people have a relatively easy time managing their behavior and achieving their goals, while other people struggle greatly. Variation in self-regulation can be explained as a function of individual differences (Brown et al., 1999; Hoyle & Davisson, 2016; Tangney et al., 2004). For example, personality traits like conscientiousness and impulsivity, which are generally stable during adulthood (Jackson et al., 2010), capture variance in behaviors and outcomes that reflect skill in self-regulation. Further, skill in self-regulation can be conceptualized and expressly measured *as* an individual difference, with items that ask people to report directly on their typical success in goal pursuit (e.g., as is the case with Trait Self-Control, Tangney et al., 2004). In addition, basic cognitive skills – like how quickly and effectively people can inhibit prepotent responses in controlled laboratory tasks – explain variation in effective self-regulation (Fujita, 2011), although to a much smaller degree than previously thought (Eisenberg et al., 2019).

Success in self-regulation is also influenced by qualities that vary at the level of a goal, including properties of goals themselves and properties of the broader context of pursuit (Carver & Scheier, 2011; Kruglanski et al., 2002). Unlike with individual differences, effects associated with goal-varying properties can be both observed (Avishai et al., 2019; Emmons & Diener, 1986) and experimentally manipulated, often in laboratory settings, which has allowed researchers to establish causal relationships between goal-varying properties and goal outcomes, like performance and motivation. Experimental approaches have found that variation in goal pursuit outcomes can be caused by, for example, how difficult a goal is (Latham & Locke, 1991), how specifically the objective of a goal is articulated (Wright & Kacmar, 1994), how close one is to attainment (Heath et al., 1999), and even how a goal is verbally phrased (Bryan et al., 2011). Further, goal-varying factors can interact in their influence on goal pursuit (e.g., Koo & Fishbach, 2008; Köpetz et al., 2011).

Individual differences and goal-varying factors operate together, in potentially complex ways. People skilled in self-regulation seem to set and approach goals wisely in ways that confer motivational advantages (e.g., Converse et al., 2019). The precise causal mechanisms linking people's skills and variable qualities of goals and environments can be difficult to parse and are bi-directional if not teleological. Skill in self-regulation is expressed and measured as skill managing dynamic properties of goals and pursuit. Further, tendencies to leverage goal-varying properties strategically are indicators of self-regulation. For example, the act of goal setting itself is used to indicate self-regulatory skill in self-report measures ("When I want to achieve something, I set goals," Ludwig et al., 2018).

Goal-varying factors covary with individual differences and also interact with them. For example, goal importance is a central component of motivation and on average, people are more motivated to pursue and achieve goals that are important. However, perceived goal importance can be *de*-motivating to people low in trait self-control (Davis & Haws, 2017). Hundreds of laboratory studies have documented moderation of goal factors by people's traits, including those not obviously related to self-regulation. For example, self-esteem is only weakly associated with measures that directly capture skill in self-regulation, but it moderates effects of some goal-varying properties in laboratory contexts (e.g., Di Paula & Campbell, 2002; McFarlin et al., 1984).

1.2 Leveraging research findings to help people in their goal pursuit

Individual differences and goal properties not only explain variation in self-regulation, they also offer insight into how people can more effectively pursue their goals. Efforts to directly improve individual differences in self-regulation have been unsuccessful, despite great interest and investment in research exploring the possibility. For example, training games and programs focused on basic cognitive skills (e.g., inhibitory control) do not produce consistent and meaningful changes in people's day-to-day self-regulation (e.g., Chen, 2016). No study has produced robust evidence in support of training programs, for example, by using a large sample, showing an unconditional effect or a preregistered conditional effect, or by showing a consistent effect across outcomes. Practice, like that offered by "brain-training" programs can improve performance on specific cognitive tasks, but those improvements do not transfer to more meaningful goals and behaviors (Simons et al., 2016).

People can't directly improve their trait skill in self-regulation through training, but they can learn strategies that may improve their goal outcomes. Indeed, a cottage industry has been built on the basis of leveraging extant theory and research on beneficial goal-varying factors. Countless popular press books and articles written by both scholars and consumers of research advise people on ways they can be more efficient and successful in their goal pursuit. Often, advice is domain-general; the recommended strategies are for goal pursuit, broadly construed, and pertain to different theoretical processes, like goal setting (Locke et al., 1981), anticipating and selecting situations that support pursuit (Duckworth et al., 2016), committing to pursuit in ways that make follow-through more likely (Rogers et al., 2014), and monitoring progress in ways that bolster and sustain pursuit (Koo & Fishbach, 2012).

However, as I argue in the next section, prior research doesn't support domain-general, broad claims about the effects of goal-varying factors on ordinary goal pursuit. Broad generalization is rarely tested, but when it is, for example, when laboratory interventions are subjected to large-scale randomized controlled trials, it is often the case that effects of goal-varying properties on goal outcomes are non-replicable, or replicable but more heterogeneous and sometimes less practically meaningful than previously understood (Gravert & Olsson Collentine, 2019; Kristal & Whillans, 2020; Miller, 2019; Oreopoulos & Petronijevic, 2019; Yeager et al., 2019). The very broad, universal advice commonly offered to people, therefore, is not yet credible. More real-world testing and more evidence from descriptive and prediction-focused approaches are needed to complement prior research so that it can yield credible advice for goal pursuit.

1.3 Limitations of prior research

Broad, universal advice about goal-varying factors is not warranted given limitations of prior research. These limitations relate to an excessive focus on developing and elaborating on theories. At scale, this tendency has yielded a body of research and theories about goal pursuit that are of limited use in understanding and explaining goal pursuit in daily life, let alone identifying universal causes of success and universal strategies for success. One exception is research on individual differences in self-regulation, which uses survey methodologies that are in many cases useful for understanding dynamics at work in ordinary goal pursuit.

In research on goal-varying factors (rather than individual differences), there are two key limitations that prevent direct application of prior theoretical research to ordinary goal pursuit. First, most claims and theories rely on evidence that is likely irrelevant to ordinary goal pursuit because it comes from settings that do not correspond to daily life. Most research on the effects of goal-varying properties come from laboratory paradigms that lack construct validity or from experiments conducted in naturalistic contexts that, though more ecologically valid than laboratory studies, still often fail to instantiate many qualities of ordinary goal pursuit. Second, many claims and theories are unsuited to application in domain-general, ordinary contexts because they involve constructs that themselves do not generalize (e.g., constructs that can only be validly operationalized in some paradigms or settings). In the following sections, I elaborate on these limitations before discussing the value of prior research and theory despite these limitations.

1.3.1 Lack of construct validity with respect to ordinary goal pursuit

Theoretical insights about self-regulation and goal pursuit have largely come from assessing people's performance on isolated tasks in unnatural circumstances that manifest theoretically important aspects of goal pursuit and self-regulation. When developing and testing theory, internal validity and careful isolation of causal mechanisms that explain the focal effect are, rightfully, prioritized. For this reason and because of feasibility and practical limitations on resources, even very broad theories are tested in narrow, often contrived, contexts where factors can be isolated and manipulated, but where it is not always possible to determine how a phenomenon actually occurs in normal settings or to even approximate the true effect size of a phenomenon.

Much prior research is limited by a reliance on evidence from laboratory paradigms. Behavioral expressions of self-regulation in the lab are often elicited with executive function tasks because basic executive functions, like the capacity to inhibit automatic responses, are thought to underlie self-regulatory behavior in daily life, like resisting temptations to smoke or eat unhealthy food (e.g., Houben et al., 2011). Simple executive function tasks are thus used to index broader self-regulatory skill. For example, in the Simon task, people must quickly press one of two buttons (red with their left hand or blue with their right hand) depending on the color of a dot that appears on the left or right side of a screen. When the locations of the dot and the correct button are aligned (when the red dot is on the left side), pressing the correct button is easy. When trials are incongruent (when the red dot is on the right side, but the button people should press is on the left), pressing the correct button requires that people override an automatic response to match right with right and left with left. How well and how quickly people perform on the incongruent trials in the Simon task is thought to measure people's basic capacity for self-regulation (e.g., as in Mani et al., 2013).

Despite the intuitive logic that studying behavior in a simplified, controlled context offers an objective window into self-regulation, simple laboratory tasks are poor measures of everyday self-regulatory skill. Performance on executive function tasks like the Simon task is not reliably associated with real-world goal pursuit behavior and tendencies (Eisenberg et al., 2019; Saunders et al., 2018). The same is likely true of even more goal-like laboratory tasks. For example, how long people work on impossible anagrams is a very popular measure of success (and persistence) in goal pursuit (e.g., Koo & Fishbach, 2012), but there is little evidence that this measure relates to actual success or persistence in personal goals. The most robust evidence to date suggests that how long people persist on an impossible anagram is uncorrelated with reliable measures of trait differences in self-regulatory skill (e.g., conscientiousness; Ebersole et al., 2016). Other studies use proxies for real-world goal-related choices, like having participants choose between ice cream and apples (Stillman et al., 2017). It is not known how similar the causal factors that produce such behavior in the lab are to those that produce meaningful goal pursuit outcomes in daily life.

Theories of self-regulation and goal pursuit are almost never exclusively developed in laboratory settings, however. Theoretical research often incorporates findings from field experiments or other naturalistic research contexts. Yet these approaches, too, often do not operationalize goal pursuit as it occurs naturally in people's lives. Many naturalistic experiments focus on factors at work during individual episodes of goal pursuit that are started and finished in one continuous episode that last minutes or hours (e.g., running a marathon), rather than more complex episodic pursuits that involve stopping and starting and last weeks or longer (e.g., training for a marathon). Many defining features of ordinary goal pursuit can't be easily simulated in research settings, like those related to the passage of time (Etkin, 2019), the interpersonal context of pursuit (Fitzsimons et al., 2015), and resource constraints that limit people's ability to pursue personal goals.

Laboratory paradigms and constrained naturalistic experiments can in many circumstances play a critical role in scientific discovery and reasoning. These approaches enable the isolation of causal mechanisms and careful theory refinement. It is problematic, however, if experimental approaches do not use operationalizations that correspond to ordinary goal pursuit. Many operationalizations used to study goal-varying factors don't resemble and are of limited or unknown empirical relation to real-world self-regulatory processes. Further, some portion of experimental work on goal-varying factors has qualities – an improbable number of positive findings given low statistical power (Schimmack, 2012), distributions of *p*-values that are not right-skewed (Simonsohn et al., 2014), and "uncanny mountains" of *p*-values between .01 and .10 in multi-study packets (Rohrer, 2018) – that suggest research practices now known to result in findings that do not replicate (John et al., 2012; Simmons et al., 2011). Thus, approaches to studying goal-varying factors in the laboratory and in naturalistic settings have generally yielded evidence that is of limited relevance to ordinary goal pursuit.

1.3.2 Theoretical constructs that are domain- or context-specific

A second way that prior research on goal pursuit is of limited use for understanding and explaining variation in ordinary goal pursuit is that many theories involve constructs that might not have meaning, or might have many different meanings, in the messy reality of daily life. Goal pursuit in daily life is extremely varied. Relatively little research on goal pursuit has been conducted on naturalistic, domain-general goal pursuit, where participants pursue idiosyncratic goals (but see Converse et al., 2019; Hofmann et al., 2012; Milyavskaya & Inzlicht, 2017; Veilleux et al., 2018). Instead, many theories are developed and refined in the context of constrained settings and involve theoretical constructs that are defined with respect to a particular domain, paradigm, or research context and cannot be easily scaled to the complexities and heterogeneity of all ordinary goal pursuits.

For example, goal pursuit outcomes are difficult to validly define across goal types and pursuit contexts. For example, although the construct of success is relatively easy to operationalize and conceive of in the context of a laboratory goal like completing a set of puzzles, there is no universal notion of success in ordinary goal pursuit. Literal, objective achievement is not always success. For example, people can primarily derive value from the act of pursuit itself rather than from what pursuit accomplishes (Freund & Hennecke, 2015). Or, in other cases, objective achievement doesn't match the goal type. Some goals are maintenance goals and do not involve achievement or attainment (Etkin, 2019). Further, because people have many different goals, with some that are more important than others, success in ordinary goal pursuit might be best defined holistically (e.g., with respect to all active goals) rather than with respect to one goal. Similar challenges for measurement arise for many other theoretical goal outcome constructs in daily life like goal performance, goal progress, and goal disengagement.

Many theoretical constructs can be validly operationalized outside of the laboratory with some adjustment or consideration, but others might evade valid measurement outside of the lab entirely. For example, how realistic or attainable a goal is can be operationalized in the laboratory (e.g., with impossible puzzles), but it can be neither manipulated nor validly measured in naturalistic contexts (e.g., as discussed by Avishai et al., 2019). In people's lives, there is no ground truth of what they can or cannot accomplish. Thus, in tests of theories that rely on the attainability construct, researchers must manipulate or measure something else, like people's beliefs about how realistic a goal is or what they can and cannot accomplish (e.g., as in Avishai et al., 2019). In individual research projects, using close proxies to measure important and otherwise unmeasurable theoretical constructs is a pragmatic solution that enables conducting ecologically valid research. However, for entire bodies of research supporting a theory, it creates jingle-jangle fallacies and can mask a critical issue: that a theory doesn't make much sense in naturalistic settings.

Theoretical constructs that do not have clear meaning in ordinary goal pursuit (e.g., "success") limit the relevance of prior research and theories, complicating hypothesis testing in naturalistic settings. Constructs that can't be validly measured in meaningful settings muddles inferences and slows the rate of knowledge accumulation. Further, constructs that can't be validly measured in daily life pose a greater scientific threat: they render theories unfalsifiable (i.e., if it's not possible to measure attainability validly in ordinary goal pursuit, no theory of domain-general context-general goal pursuit involving that construct can ever be disproven). Therefore, issues of construct validity with respect to ordinary goal pursuit can precluding meaningful and conclusive hypothesis testing. Thus, because many extant theories about goal-varying properties rely on constructs that do not themselves generalize to ordinary life, they are of limited use, and may not be usable at all, in understanding ordinary goal pursuit.

1.3.3 Consequences for self-regulation research

An ironic consequence of focusing on theory over prediction and description is that our evidence and theories may not be useful in helping us understand the causal processes that give rise to the outcomes we care most about (i.e., those that occur in the real world). Due to a prioritization of theoretical inference over description and prediction, prior research on goal-varying factors has generally taken place in constrained settings. Theories and effects have not been systematically tested across samples, goals, and contexts that represent real-world variation and thus does not offer direct evidence for broad generalizability. Further, although direct evidence for broad generalizability is not necessary when there is strong theory, theories of and related to goal pursuit are poorly suited to ordinary pursuit contexts. Theories were largely developed in constrained settings and sometimes rely on constructs that don't scale to everyday contexts.

What value is there in extant theory, given these limitations? This is an empirical question, and one that likely has many answers and depends on specific theories and specific real-world contexts. Still, it is unlikely that all extant theories are true in their broadest form. Theories of goal pursuit are often simple, unqualified verbal theories of ordinal predictions (e.g., "X causes Y to increase") rather than qualified and precise mathematical ones (e.g., "$Y = .32X^2$ under conditions A, B, and C"). These broad and simple, often linear, theories are unlikely to describe the actual causal process at work (Yarkoni & Westfall, 2017). For example, success (or failure) in a goal like smoking cessation, is the outcome of thousands and thousands of individual instances of behavior which each have complex causes, many of which are not purely psychological, and many of which are unlikely to be governed by linear or even polynomial functions. Self-regulation is a complex system and our models and theories likely lack appropriate complexity to cover all kinds of goals. Other areas of psychological science that pertain to very complex phenomena have identified the need for theories that describe non-linear, dynamic relationships (e.g., the use of chaos theory in the study of romantic relationships;

Weigel & Murray, 2000). In the absence of research focused on description, characterization, and prediction of real-world outcomes, the accuracy and practical utility of our theories and theoretical models is unknown.

However, even if theories of goal pursuit and self-regulation do not function as precise models of causal processes and even if they cannot be used to accurately predict outcomes of people's smoking cessation or other goal pursuits, they can still have value in identifying goal-varying factors that have potentially causal relationships with goal outcomes. Extant theories in goal pursuit and self-regulation research are based on decades of careful thought and integrate information from people's observations, lived experiences and clinical expertise, as well as from individual differences research that asks about people's tendencies in goal pursuit. These sources of information *do* reflect the complexity of daily life and the heterogeneity of ordinary goal pursuits. Thus, while it is unlikely that any general theories of goal pursuit can predict outcomes with specificity, goal-varying factors identified as important by theories likely can explain variance in real-world outcomes and can contribute to accurate predictive models.

Further, a growing number of empirical studies are not subject to the limitations discussed above. Theory-informed descriptive research is an emerging trend in self-regulation research and theory-testing is increasingly done in ecologically valid settings with meaningful outcomes (e.g., Wilkowski & Ferguson, 2016). In addition, there is a long history of describing naturalistic goal pursuit in personality psychology; the content of people's goals (i.e., their "personal strivings," "motives", and "values") have long been conceived of as a dimension of their personality (Emmons, 1986). These approaches to research have yielded information about ordinary goal pursuit, like people's daily experiences of temptations (Hofmann et al., 2012; Veilleux et al., 2018), people's desires during the process of goal pursuit (Converse et al., 2019), and the skills and strategies associated with real-world goal pursuit outcomes (Hennecke et al., 2018; Ludwig et al., 2018). These efforts demonstrate the value of descriptive research and research situated in uncontrolled settings. Work that is unconstrained by existing theory has the potential to make new discoveries, like previously undocumented kinds of goal outcomes (e.g., Davydenko et al., 2019), or barriers to goal pursuit that don't relate to motivation or other influential theories of goal pursuit (e.g., failing to remember plans or intentions for pursuit; Einstein & McDaniel, 2005).

1.3 Advancing self-regulation research

Hundreds of research studies pertain to self-regulation and goal pursuit, and dozens of theories describe potential mechanisms of success in goal pursuit. Two main mechanisms of success are people's individual differences in self-regulatory skill and goal-varying factors that help pursuit. However, due to limitations of prior research on goal-varying factors, there are critical gaps in our understanding of mechanisms of success in *ordinary* goal pursuit (compared to goal pursuit as it occurs in more constrained settings). In addition, little is known about how individual differences and goal-varying factors covary and interact during the course of ordinary goal pursuit. Despite this relative lack of evidence, claims of universal effects of goal-varying properties on goal pursuit are prevalent, and imply that small adjustments to how a goal is specified or thought of can make the difference between success and failure. Such claims of universal effects across all goals are very common, and so it's easy to overlook how extraordinary these claims are and how large and robust an effect would have to be to meaningfully affect goal pursuit regardless of goal content and people's particular skills and contexts.

This dissertation is designed to complement theoretical research on goal-varying factors that promote success in goal pursuit, to advance our understanding of the factors that affect goal pursuit in daily life. It has two main aims. First, this dissertation aims to *characterize* one kind of ordinary goal pursuit – New Year's resolutions. Second, it aims to *identify goal-varying factors* that longitudinally predict subjective success and to estimate how well goal-varying factors can predict people's ratings of their goal achievement. The studies make use of extant theories to identify a handful of goal-varying factors that may explain and predict variance in goal outcomes over and above an individual difference measure of self-control.

This dissertation accomplishes these aims by adopting methodological and analytic approaches that are uncommon in social psychological research on goal pursuit. I take a longitudinal survey approach in a sample of American adults as a pragmatic and informative method for conceptually replicating prior findings about the influence of nine goal-varying factors on goal pursuit outcomes. This allows me to also assess and preserve the natural covariation of goal-varying factors among typical goals. Further, because people report on multiple New Year's resolutions, I can explore effects of individual differences separately from effects of goal-varying factors. Many theories of goal processes relate to within-person (goal-level) processes, but the data used to develop and test these theories often compare people, not goals within people. Processes that operate between people are not necessarily the same as those that operate within people, and processes within people might differ from person to person. Thus, considering between- and within-person effects separately can yield a more nuanced understanding of how goal-varying factors might benefit goal pursuit.

In addition, I use analytic “machine learning” tools developed in computer science, including cross-validation. These tools allow me to assess the predictive value of ordinary regression analyses, and to conduct robust exploratory analyses that focus on optimizing prediction. As others have noted, tools from machine learning are a natural fit for psychological science, particularly when our interest is in predicting meaningful outcomes that have complex causes (Dwyer et al., 2018; Yarkoni & Westfall, 2017).

In Chapter 2, I review prior naturalistic and observational research on goal pursuit and New Year’s resolutions, and present Study 1. The aim of Study 1 is to characterize ordinary goal pursuit, and to identify potential factors that are associated with success longitudinally using analytic approaches that are typical of social and personality psychology. In Chapter 3, I present Study 2. The aim of Study 2 is to identify factors that robustly and longitudinally predict goal outcomes using machine learning analytic approaches. Study 2 also offers an opportunity to replicate findings from Study 1. In the final chapter, I summarize the findings, discuss their limitations, and broadly discuss the current status and potential future directions for the study of self-regulation and goal pursuit.

2. Study 1: Describing New Year's resolutions

2.1 Introduction

This study aims to characterize people's pursuit of New Year's resolutions, focusing on basic description of goals, pursuit, and goal outcomes as well as goal-varying factors that theories identify as important to success. In addition, this study explores associations among goal-varying factors, between goal-varying factors and a measure of skill in self-regulation (trait self-control), and between goal-varying factors and subjective success in pursuit, accounting for trait self-control. The analyses are organized around eight research questions, as motivated and summarized in the next sections.

2.1.1 What goals do people set as New Year's resolutions?

Goals are cognitive representations of desired (or undesired) future states (Elliot & Fryer, 2008). The goal construct encompasses cognitive representations that vary in concreteness, level of intention, difficulty, time range, deliberateness, and complexity (Austin & Vancouver, 1996). Numerous theoretical models and hierarchies describe the structure and content of everyday goals (Austin & Vancouver, 1996; Kruglanski et al., 2002; McAdams, 1996). Typically, goals are thought of as hierarchically organized, with abstract values or motives guiding major life goals (Roberts et al., 2004), which guide more specific mid-level goals, like personal strivings or personal projects (Emmons, 1986; Little et al., 1992), which in turn guide more concrete action relevant in the immediate future.

New Year's resolutions likely fall into the mid-level goal category based on their scope and typical content. The same basic themes tend to emerge in people's mid-level goals across different language and solicitations; When asked to describe personal strivings, wishes, personal projects, New Year's resolutions, and personal goals, responses are broadly similar and relate directly to concrete domains like relationships, career or education, finances, community, health, and religion or spirituality (King & Broyles, 1997; Reisz et al., 2013; Salmela-Aro et al., 2012; Woolley & Fishbach, 2016). Consistent with a coherent hierarchy of goals, the basic content of people's mid-level goals relates to their personality (e.g., Emmons, 1986) and other broad dispositions (e.g., Goodman et al., 2019).

2.1.1.1 Content of resolutions

Prior studies of New Year's resolutions measured resolutions with broad self-report nomothetic categories rather than open-ended text (e.g., Woolley & Fishbach, 2016). However, prior studies of mid-level goals, and in particular people's personal strivings, have used idiographic approaches combined with nomothetic coding based on existing manuals or by bottom-up categories, derived qualitatively or using automated methods (McAuliffe et al., 2020; Veilleux et al., 2018).

Although nomothetic coding demonstrates that New Year's resolutions are broadly similar in their content to other mid-level goals (i.e., at the level of domain), whether and how New Year's resolutions differ from other mid-level goals isn't exactly known. It may be the case that resolutions are just a sample of ordinary goals. People may set as resolutions the goals they had intended to pursue, inspired either by social tradition or because of the new year itself (Hennecke & Converse, 2017). However, New Year's resolutions are not ordinary goals and may have unique characteristics. Even if they are ordinary mid-level goals, their timing alone might shape their content. For example, even though health goals are generally prevalent (e.g., in a sample of 557 American Mturk workers, 66% had health goals; Milyavskaya & Nadolny, 2018), New Year's resolutions may be more likely to relate to health goals than goals set at other times of the year. Many people celebrate cultural and religious holidays at the end of the year that for many involve indulgent eating and travel, which may bring lapses in dietary and exercise routines.

Research question 1: What goals do people set as resolutions?

2.1.1.2 Properties of resolutions

2.1.1.2.1 Concreteness and specificity

Prior research suggests conflicting possibilities about the specificity and concreteness of New Year's resolutions. Goal specificity and concreteness are related but distinct constructs that have each long been identified as motivationally important properties of goals (Heath et al., 1999; Höchli et al., 2018; Locke et al., 1981; Wallace & Etkin, 2018). Goal specificity refers to how precise a goal's endpoint or objective is. Goal concreteness or abstractness refers to where a goal is in the goal hierarchy, which has abstract and general goals at the top and concrete and narrow goals at the bottom. Specific goals with clear endpoints and concrete goals that would manifest in observable or measurable outcomes benefit self-regulation because they can be tracked and monitored, and so people can get feedback on their pursuit and adjust as needed. There are discrepant views on the effects of abstract versus concrete goals on self-regulation and aspects of goal pursuit, like goal progress (Fujita et al., 2006; Höchli et al., 2018;

Trope & Liberman, 2010; Vallacher & Wegner, 1987). These discrepant theoretical perspectives often use different operationalizations and paradigms, and are typically conducted in laboratory settings, which makes it difficult to adjudicate and determine a directional hypothesis in the context of ordinary goal pursuit.

There is little work on how concrete and specific goals typically are, and so it is unclear how concrete and specific New Year's resolutions are. On one hand, resolutions might be vague; they might relate to abstract constructs that can't be observed directly and thus resolutions may involve end points that can't be specifically articulated. Temporal landmarks, like the beginning of a new year, are associated with broad aspirations (Dai et al., 2014). Thus, resolutions may refer to broad values and purposes.

On the other hand, resolutions might be concrete and specific; they might relate to observable constructs that can be measured and articulated specifically. Resolutions are implicitly time-constrained, and so they may be more specific and concrete than other kinds of mid-level goals. For example, resolutions might be more concrete and specific than personal strivings which are chronic rather than of-the-moment goals (i.e., the elicitation of personal strivings asks people to list the things they characteristically try to do; Emmons, 1989). Thus, someone might have a personal striving to "take care of my body" which is expressed as concrete resolution to "do more weight training" or even more specifically to "lift weights weekly."

Research question 1a: How concrete are New Year's resolutions?

Research question 1b: How specific are New Year's resolutions?

2.1.1.2.2 Approach, avoidance, and maintenance

A classic distinction in self-regulation theories is that of approach and avoidance. The distinction between approach and avoidance is relevant to many different theoretical perspectives and units of analysis: People have basic tendencies related to approach and avoidance in goal pursuit (Elliot & Thrash, 2002), people can adopt approach and avoidance motivation for a goal (Elliot & Harackiewicz, 1996), and goals themselves can be conceived of as approaching a desired state or avoiding an undesired state (e.g., Hennecke, 2019). Goals can also focus on maintaining a current state or can combine different targets or ranges of targets (Brodscholl et al., 2007; Wallace & Etkin, 2018). Numerous theories describe why and how approach, avoidance, and maintenance goals can affect motivation and performance in a broad range of domains (Albarracín et al., 2018; Brodscholl et al., 2007; Darnon et al., 2007; Elliot & Harackiewicz, 1996; Hennecke, 2019).

Whether New Year's resolutions are more typically one kind of goal is unclear from prior literature and may be difficult to conclusively assess. The theoretical construct of interest is how people think of their goals (i.e., people's conceptualization of a goal as approach, avoidance, maintenance, or some combination), but in goals that span weeks and months, people may think of the same goal at different times as relating to approaching a desired state, avoiding an undesired state, maintaining a current state, or a combination of those. Indeed, there is likely a strategic benefit to doing so. Dynamically shifting focus on approach or avoidance as best fits ones' current goal pursuit needs allows people to leverage the motivational properties of these conceptualizations (e.g., Hennecke et al., 2018).

Given the theoretical legacy of approach and avoidance properties and the likelihood that, so long as they can be validly measured, they relate to achievement, assessing them in the context of New Year's resolutions is worthwhile.

Research question 1c: What proportion of resolutions are focused on approaching a desired state, avoiding an undesired state, and maintaining a current state?

2.1.2 How motivated are people in their New Year's resolutions?

Another unique feature of New Year's resolutions, relative to other kinds of ordinary goals, is that they are not spontaneous and so they may have a different motivational profile than other goals. People may set New Year's resolutions out of a sense of tradition rather than because of a genuine desire for change. For this reason, people might not have as much motivation (or may have short-lived motivation) for their New Year's resolutions relative to other goals they set and pursue. Motivation can not only have direct effects on achievement, it may also moderate effects of other factors (e.g., goal properties, Locke, 1968; but see Hollenbeck & Klein, 1987).

Motivation is the force that drives intentional goal pursuit and is a key theoretical construct in understanding success and related goal pursuit outcomes, like persistence. Motivation is a leading indicator of success. It originates in the reasons that people are engaged in pursuit (e.g., whether autonomous or controlled, Deci & Ryan, 2000) and in the net value people expect to derive from the goal (or its pursuit) and the likelihood of achieving it (i.e., motivation reflects expectancy, positive value, and negative value or cost; Atkinson, 1957; Flake et al., 2015; Klinger, 1975; Wigfield & Eccles, 2000). In choice models of self-regulation, motivation is encapsulated by the construct of subjective value which reflects the totality of motivational forces in a moment, including positive rewards and negative costs. Subjective value (and motivation) also reflects people's contextualized and dynamic evaluations and experiences over the course of goal pursuit (Berkman et al., 2017; McGuire & Kable, 2013). Because of these complexities, motivation not only shapes success, it is also responsive to it; the more motivated people are, the more they work on their goals, which increases their motivation.

In the study of self-regulation and goal pursuit, motivation has no standard measurement, and is often measured with self-report items that correspond to various aspects of motivation including its components (e.g., confidence, efficacy, value), its consequences (e.g., effort, persistence, intention to continue pursuit), or with items designed to more or less directly capture it (e.g., commitment, autonomous or controlled motives).

Research question 2: How motivated are people in their pursuit of New Year's resolutions in terms of their confidence, effort, and commitment at the beginning of the year? How does motivation change throughout the year?

2.1.3 How do people pursue New Year's resolutions?

New Year's resolutions are likely quite heterogeneous in their content and properties, and how people pursue their resolutions are likely quite varied, too. Pursuing a goal is not just a matter of making progress on the goal; it involves many other self-regulatory behaviors. The classic model of self-regulation characterizes it as a control system (Carver & Scheier, 1982). In this view, people pursue and achieve goals much like a thermostat controls a room's temperature: by routinely assessing their status and adjusting their goal-directed behavior based on the discrepancy between where they currently are and where they want or expect to be with respect to the goal. Contemporary process models of self-regulation elaborate on this simple control system to include additional processes involved in controlling thoughts, behavior, and emotions in service of goals (e.g., as described in Hoyle & Gallagher, 2015). No prior study of New Year's resolutions has examined the behaviors people associate with pursuit, assessed the costs that pursuit entails, or catalogued the spontaneous use of effective strategies in the context of specific goals.

Research question 3: How do people pursue their resolutions?

2.1.3.1 Social commitment

Using commitment devices, like making Social Commitments related to goal pursuit, is an effective strategy because it creates costs (i.e., consequences) for stopping pursuit. For example, agreeing to contracts that result in the loss of money if people fail to continue pursuing their goals is effective in helping people stick with their goals (for a review, see Rogers et al., 2014). The potential costs of an effective commitment device can also be reputational and not just financial. Making a commitment to a friend and letting other people know about intentions is an effective self-regulatory strategy. Most studies of commitment devices have been conducted in field settings using randomized control trails. However, little is known about spontaneous use of Social Commitment in ordinary goal pursuit, even the most simple form, of telling others about ones' goals.

Research question 3a: How often do others know about people's New Year's resolutions?

2.1.3.2 Habit formation

Another effective pursuit strategy that people can use is Habit Formation. Habits supports many behaviors in everyday life (e.g., locking a car door, making coffee, looking at mobile devices) and are characterized by automaticity, or daily or near daily engagement, independence of intention and behavior, independence of goal-directed behavior and goal-relevant thought, limited emotional responses associated with the behavior, and stability of behavioral context (Wood et al., 2002). These characteristics make habits an asset in goal pursuit. When a behavior is habitual, pursuit is no longer contingent on deliberate intention, and is protected from amotivation and self-control conflicts (Carden & Wood, 2018; Neal et al., 2013; Wood et al., 2002).

Habits form when people engage in a reinforced behavior regularly, and in a stable context (Wood & Rünger, 2016). They can be an asset to intentional behavior change; If people can engage in effortful goal pursuit consistently during the initial weeks of pursuit, habits theoretically take over. In one study, goal-supportive behaviors that begin as deliberate and effortful reached a stable point of automaticity – that is, they became habits – after an average of 60 days (range: 18 to 254, in a model that fit 39 of 62 total participants; Lally et al., 2010).

Although extant evidence is compelling, many basic questions remain about the role that habits play in ordinary goal pursuit. At a minimum, there seems to be an association between Habit Formation in everyday life and success: People who tend to be good at achieving their goals often form habits (Galla & Duckworth, 2015). However, how habits form in the messy reality of daily life and how stable they are is not well understood. Further, habits may not be appropriate for all pursuits. For some goals, pursuit is necessarily sporadic or necessarily occurs in variable contexts. Further, some people may not have the basic stability that habits require; for example, if they move homes often or their working schedule is variable.

Research question 3b: To what extent do people form habits in their pursuit of New Year's resolutions?

2.1.4 *How successful are people in their* New Year's *resolutions?*

Only one known prior study has prospectively tracked goal pursuit of New Year's resolutions over the course of at least one year. Norcross and Vangarelli (1988) administered phone interviews in January to 200 people in the United States who set New Year's resolutions and followed up after one week, two weeks, three weeks, one month, three months, six months, and two years. Success, measured on a four-point Likert scale (1 = *totally failed*, 2 = *mostly failed* , 3 = *mostly succeeded*, 4 = *totally succeeded*) was initially high but steadily dropped over the course of the study: 60% of people reported success at three weeks, 43% reported success at three months, and just 19% reported success at two years.

Other prospective longitudinal studies, focused on much shorter time intervals, provide additional support for the idea that people start the year feeling strong in their resolutions, but success tapers off as the year progresses. Woolley and Fishbach (2016 and Supplemental File) found high rates of average self-reported persistence among

Mturk workers ("…how successful have you been at sticking with this resolution?", 1 = *not very successful*, 7 = *very successful*) after three weeks (M = 4.73, SD = 1.70, N = 101) and two months of working towards New Year's resolutions (M = 4.79, SD = 1.56, N = 242). However, around three months, success may taper off (Höchli et al., 2019). In a sample of 365 online participants, average success rates at three months (1 = *unsuccessful*, 7 = *successful*) were below the midpoint (M = 3.60, N = 256; calculated with information provided in the Supplemental File). In addition, a surprising number of (responding) participants had achieved their resolution after just three months: nearly 18% (46 of 256). This high rate of early achievement in this study is likely due to the encouraging reminders participants received every two weeks over the course of the study.

Finally, a few retrospective surveys suggest that many people set resolutions and report success in achieving at least some of them. Surveys of adults in Great Britain from the years 2015 (Bupa, 2015; N = 1937), 2017 (YouGov, 2017, N = 1629), and 2019 (YouGov, 2019; N = 2020), estimated that 12%-25% of adults set resolutions and about half achieve at least some of their resolutions (47%-57%) with 24% in 2019 reporting that they achieved all of their resolutions. Retrospective reports may be inaccurate, however, because they require that people remember what resolutions they set but gave up on. Indeed, between 4% and 11% of people surveyed reported that they didn't know or couldn't remember when asked if they had set a resolution the previous year.

A challenge of naturalistic approaches to studying goal pursuit is that success can be difficult to validly operationalize. Evaluations of personal goal pursuit are inherently subjective, and any objective outcomes are qualified by unknown context. Even setting the issue of subjectivity aside, there is measurement invariance in assessing New Year's resolutions. Whether a resolution was literally achieved by the end of the year or how much progress was made towards the resolution has a different meaning (i.e., would load differently on a ”success” factor) depending on, for example, whether a goal had a defined, measurable end-state (e.g., run a 5k), or not (e.g., be healthy). People also modify goals during pursuit, which can introduce ambiguity about the measurement and meaning of success. Given this, characterizing success in an observational survey of New Year's resolutions requires the use of multiple measures of success, including measures of objectively achievement and Subjective Success.

Research question 4: What proportion of people achieve their New Year's resolutions? How do different measures of success (e.g., goal status, literal achievement, and Subjective Success) compare?

2.1.5 How often do people disengage from their New Year's resolutions?

Persistence is essential to accomplishing our most important and meaningful personal and social goals, but people must sometimes quit one goal for the sake of others. Ironically, giving up one goal can optimize broader success and well-being in the long-term. For example, when people are facing impossible goals, people who give up on them more easily experience less physical and psychological distress (Wrosch et al., 2013).

The decision to quit or persist is complex. It may be adaptive to quit impossible goals, but in everyday life people rarely know which goals are impossible and which are merely difficult. Navigating these choices well and giving up on goals when it is beneficial to do so is theorized to be an essential self-regulatory skill (Wrosch et al., 2003). Theories describe disengagement as deliberate and involving weighing the costs and benefits of continuing versus giving up (Brandstätter & Schüler, 2013). Although aspects of this deliberation process and the impulse to disengage have been studied in mid-level goals (Brandstatter et al., 2013; Brandstätter & Schüler, 2013), very few studies assess actual disengagement in long-term mid-level goals (but see Herrmann & Brandstätter, 2015). Much remains to be learned about disengagement as it occurs in ordinary goal pursuit.

Research question 5: How often do people disengage from their resolutions? How often do people think about disengaging from their resolutions?

2.1.6 Which goal-varying factors tend to coincide?

This dissertation focuses on factors that are individually theorized to predict success, but in ordinary goal pursuit, these factors may coincide. Few prior studies have observed multiple goal-varying factors simultaneously, which has precluded assessment of ordinary covariation of goal-varying factors. The extent and patterns of covariation among goal-varying factors is useful descriptive information that can help accurately characterize goal pursuit in ordinary settings and it is also necessary for interpreting the meaning of associations between goal-varying factors and success in goal pursuit. To the extent that factors covary in this study, their effects are confounded. For example, if *concrete* goals are also mostly *physical* goals, then any observed effects of concreteness and physical domain status on success cannot be attributed uniquely to concreteness or to physical domain status.

Research Question 6: How do goal-varying properties covary?

2.1.7 Which goal-varying factors are associated with skill in self-regulation?

As discussed previously, skill in self-regulation often manifests in skillful selection of goals and skillful pursuit in goals. People skilled in self-regulation, such as those who score high on measures of Trait Self-Control, may pursue "better" goals and use better strategies for pursuit (Converse et al., 2019; Duckworth et al., 2016; Ludwig et al., 2018). The relationships between individual differences in self-regulation and goal-varying properties have not been systematically examined in a naturalistic context. Instead, patterns of covariation and interaction between traits and goal-varying properties have been documented in different samples and goal contexts (e.g., puzzle completion tasks and genuine ordinary goals) using different measures of individual differences and theoretical frameworks, which makes it difficult to adjudicate any broad conclusions (e.g., even with one construct and one domain; Stautz et al., 2018). To the extent that goal-varying factors are reliably associated with success, they should also be associated with Trait Self-Control.

Research question 6: How do goal-varying factors relate to Trait Self-Control?

2.1.8 What factors predict success in New Year's resolutions accounting for trait self-control?

The primary purpose of this study is to characterize and describe New Year's resolutions. However, these data offer an opportunity to explore associations between characteristics of resolutions and their ultimate success. Goal content and properties, people’s motivation, and their strategies for goal pursuit predict success in the laboratory. Here, I examine the extent to which they predict success in ordinary life. Specifically, I examine the association between Subjective Success and: goal domain, concreteness and specificity, approach and avoidance, motivation at the beginning of the year, Social Commitment, and Habit Formation. For each, I examine effects within and between people (i.e., the effects of variation within people as well as between people on Subjective Success).

Research question 6: Which goal-varying factors explain variance in Subjective Success in New Year’s resolutions, within or between people, accounting for Trait Self-Control?

2.2 Method

Four hundred fifteen people recruited through Mturk completed a survey describing up to five New Year’s resolutions (N = 1094). They also completed several individual difference measures related to self-regulatory skill and answered demographic questions about themselves. Participants were invited to complete surveys across the year, reporting on how pursuit of their resolutions was going and, ultimately, whether they accomplished their goal. Three hundred twelve participants completed a follow-up survey in April (T2), 254 participants completed a follow-up survey in July (T3), and 251 participants completed the final survey in January of the following year (T4). At the end of each survey they provided comments and reported on the validity of their data, including whether they answered questions randomly. Participants who said they answered questions randomly were compensated but their data are not included in analyses.

2.2.1 Sample

Participants were native English speakers in the United States, recruited with Amazon Mturk (M_{age} = 35.9, SD_{age} = 12.1). There were 200 male-identified people in the sample and 215 female-identified people. Participants who completed the survey were excluded from analyses if their open-ended responses were insincere or suspicious, and if they did not complete the survey and receive compensation for their work.

Participants were well-educated; 96% of the sample had a high school degree or more. The mode of educational attainment was a Bachelor's degree (38%; *n* = 158). Many people had completed some college (26%; *n* = 109) or an Associate's degree (13%, *n* = 52). About 11% of the sample had a graduate or professional degree or had completed some graduate or professional training (*n* = 45).

The sample was mostly White and indicated their race identities as follows (categories are non-exclusive): *n* = 339, 82% White; *n* = 31, 8% Black; *n* = 24, 6% Asian; *n* = 16, 4% Native American or Alaska Native; *n* = 2, <1% Hawaiian or Pacific Islander; or another identity, *n* = 9, 2%.

2.2.2 Procedure

In mid-January of 2016, participants were invited to participate in a survey, which was described as requiring that people had set New Year's resolutions. The first question participants saw was a screener that asked if they had set resolutions. If so, they were given informed consent and those who agreed to participate completed the survey. The survey, which entailed completing individual difference measures, reporting up to five resolutions, and answering questions about each one. Participants were invited to participate in additional surveys in mid-April (T2), mid-July (T3), and mid-January of 2017 (T4) as described in the initial consent form, via TurkPrime. Participants had approximately two weeks to complete each survey and were sent up to two additional invitations.

Payment for the first survey was $1 and it took participants an average of 12.5 minutes to complete. Payment for the second survey was $3 and it took participants an average of 12.4 minutes to complete. Payment for the third survey was $3 and it took participants an average of 11.7 minutes to complete. Payment for the fourth survey was $3 and it took participants an average of 15.5 minutes to complete. Participants who completed three surveys received an additional $1. Participants who completed four surveys received an additional $2.

2.2.2 Measures

Measures used in inferential analyses are described here; additional measure information related to descriptive analyses is reported in the results section. Measures derived from open-ended text are described in section 2.2.2. Many more measures were collected in this study. Surveys as administered and data are available at osf.io/qtvc6.

Motivation was measured in January (T1), April (T2), and July (T3) with three items written for this study. People reported their commitment (“Currently, how committed are you to this resolution?”), effort (“Currently, how much effort are you putting towards this resolution.”), and confidence (“How confident are you that you will achieve this resolution by December 31, 2016?"), which capture related but distinct aspects of the motivation construct. Commitment is an expression of motivation that, in Goal Systems Theory, reflects contextualized properties of the goal, including its value and available means (Kruglanski et al., 2002). Effort is a consequence of motivation, albeit a potentially noisy one that reflects barriers to pursuit in addition to latent motivation. Confidence is a component of motivation, which relates to people's efficacy (Bandura, 1982; Eccles & Wigfield, 2002). Responses were indicated on a Likert-scale from 1 (*not at all*; *none*; and *not at all*, respectively) to 5 (*very*; *a lot*; and *completely*). Reliability of the three items together (assessed with one resolution per person) was high at T1 (α =.75, ω = .77), T2 (α =.86, ω = .86), and T3 (α =.86, ω = .87). At T1, responses were skewed and likely attenuated reliability.

Social Commitment was measured in April (T2) with one item written for this study that asked participants if others knew that they had set the focal resolution ("Other people know that I have made this resolution"). Participants indicated how true of them the statement was on a Likert-scale from 1 (*not at all true*) to 5 (*completely true*).

Habit Formation was measured in April (T2) with two items written for this study but adapted from the Behaviour Frequency Context X Stability measure (BFCS; Ouellette and Wood, 1998). First, participants were asked to list the behaviors associated with working on their resolution. Then, they reported the two items that constituted the scale: "Over the last two months, how often have you performed these behaviors?" and "Over the last two months, how often have you performed these behaviors in the same place?" Participants indicated their responses on a Likert-scale from 1 (*never or almost never)* to 5 (*always*) and from 1 (*never in the same place*) to 5 (*always in the same place*). These two items were correlated with one another moderately (e.g., within people's first resolution, r = 0.35, t(282) = 6.21, p<.001).

Subjective Success in goal pursuit was measured at the end of the year (T4) with one item written for this study. Participants reported the extent to which they felt successful in a subjective sense, considering modifications to their goal and considering constraints on a Likert-scale from 1 (*not at all*) to 5 (*completely*).

Trait Self-Control was measured with the Capacity for Self-Control Scale, a 20-item measure that differs for other measures of Trait Self-Control in that it includes subscales for three varieties of self-control: inhibition, initiation, and continuation (Hoyle & Davisson, 2018). In this scale, people indicate how often their behavior reflects tendencies relating to each of these varieties of self-control on a Likert-scale from 1 (*hardly ever*) to 5 (*nearly always*). In this sample, the scale had high reliability as estimated in two different ways. First, the scale had excellent reliability as estimated by Cronbachs alpha (α =.94). Alpha underestimates internal consistency when items have different factor loadings and covariances (Revelle & Condon, 2019), as is likely the case with this three-factor scale. Reliability was also high as estimated by McDonald's total omega (ω = .95), suggesting that 95% of variance in unit-weighted scores attributable to common variance among items (Rodriguez et al., 2016). In this sample, Trait Self-Control was positively correlated with other measures of and related to self-regulatory skill: BFI-44 Conscientiousness (α =.90, r = .85; John et al., 1991); Grit (α =.89; r = .78; Duckworth et al., 2007); impulsivity (α =.81, r = 0.61; Joireman & Kuhlman, 2004;

Zuckerman, 2008); but was negatively correlated with Goal Disengagement Capacity (α =.93, r = -.34; Wrosch et al., 2013).

2.2.3 Coding open-ended text

Participants provided open-ended responses to a number of questions on the survey, including in their reports of the content of their New Year's resolutions. Open-ended text was assessed with both automated methods (i.e., word counts), and by ratings from human coders. Three independent coders rated each resolution for *concreteness* (Abstract goal tied to personal values, success is subjective; Concrete goal tied to concrete outcome measure), *specificity* (Low, goal with subjective/non-quantifiable outcome; Moderate, goal tied to an objective outcome or measured relative to current condition; High, goal with unambiguous or quantified objective outcome) and *approach* (Approach, involves moving towards a particular outcome; Avoidance, involves moving away from a particular outcome; Maintenance goals; Multiple). Substantial discrepancies were identified and resolved through discussion (initial estimates of interrater reliability were below .50 and in many cases were difficult to determine, as coders sometimes skipped resolutions that were not clearly categorizable). Final code values were determined by consensus.

2.2.4 Analytic approach

Given the number of variables and potential analyses, and the increase in false error rates associated with both analytic flexibility and large numbers of tests, I restrict inferential analyses to a relatively small set of theoretically important variables. This set of variables prioritizes those that were measured earlier in the year; identifying early predictors of subsequent success is more valuable than identifying proximal indicators of success. This set of variables also prioritizes those that were measured in Study 2, and so any effects found in Study 1 can be directly or conceptually replicated (e.g., in cases where measurement differed across studies).

2.2.4.1 Assessing covariation

Clustered data presents a challenge for identifying covariation between pairs of variables. Pearson *r* can be used to calculate zero-order and point-biserial correlations among continuous and categorial variables, but it does not account for clustering. Estimates of correlations based on all available data can result in inaccurate point estimates, depending on the underlying data structure, and is likely to result in underestimations of standard errors, which can affect inferences. To avoid deflated standard errors while optimizing interpretability of results, I compute correlations among all variables with Pearson *r*, limiting the analysis set to just the first resolution that people set. I report the covariance structure of the analysis dataset, I provide means, standard deviations, and Pearson *r* correlations derived from the entire dataset in Appendix A. Unless otherwise noted, correlations use pairwise complete observations.

To contextualize estimates of correlations from people's first resolutions only, I examine the extent to which clustering accounts for variance in each variable by calculating the Intraclass Correlation (ICC) in a random-effects ANOVA (i.e., a model that accounts for clustering of resolutions within people but is otherwise empty). ICCs estimate the proportion of variance at the resolution level attributable to people. An ICC of zero indicates no differences between people and an ICC of one indicates that all differences are between people. The higher a variable's ICC, the more similar a correlation point estimate derived from one resolution will be to an estimate of covariance derived from all available data that accounts for nesting of resolutions within people. (However, point estimates derived from one resolution per person will be more variable, and parameter tests will have less statistical power.) For variables with ICC values close to zero, estimates derived from the entire dataset will be both more precise and more accurate. Given that theories do not make different predictions for effects within people or between people, and given that there is no way to tell a priori how goal-varying factors vary between and within people, I focus on the more conservative estimate: correlations derived from people's first New Year's resolutions. This approach is more consistent with the overall aims of this dissertation in its prioritization of identifying robust effects. However, it risks missing smaller effects by reducing power.

2.2.4.1 Explaining variance in success

For inferential analyses focused on explaining variance in success, I use multilevel models where Level 1 is the resolution and Level 2 is the person. My interest is in decomposing effects attributable to resolution factors from those attributable to people, so I use group-mean centering for Level 1 predictors and models include group means as Level 2 predictors. All Level 2 predictors, including group means, are grand mean centered. This approach results in meaningful intercept estimates and parameter estimates that directly respond to within-group and between-group effects. Intercepts are the estimated value on the dependent variable of an average person's average resolution. Parameters of group mean centered predictors are the estimated within-person effect of the predictor (i.e., the effect of the predictor in comparing multiple resolutions a single person holds). Parameters of (grand mean centered) group means are the estimated between-person effect of the predictor (i.e., the effect of the predictor in comparing people with different average values across their resolutions). Some people had only one resolution, so my ability to parse within-group and between-group effects is limited by this mild confounding of between and within-person effects.

The modeling approach will follow the same process, and model defaults will be adjusted if models cannot be estimated. Following best practices in multilevel modeling, each model will be initially specified with random slopes (Barr et al., 2013). This has the additional benefit of limiting analytic flexibility. Not all models can be estimated with random slopes, however, so models will be simplified if there are convergence failures or insufficient degrees of freedom. Each model will also first be estimated using a frequentist approach to multilevel regressions (as implemented with the R package lme4; Bates et al., 2015). For models with non-normal residuals, I also estimate Bayesian multilevel models with default priors (as implemented with the brms R package; Bürkner, 2017). These models produced reasonably normally distributed residuals, which was an issue with Subjective Success due to a platykurtic (i.e., near uniform) distribution. Unlike other kinds of distributions, uniform distributions cannot be easily transformed within the context of linear models nor can they be modeled with functional forms commonly available in implementations of generalized linear models.

2.2.4.2 Other considered modeling approaches

The modeling approach I take in this study does not fully leverage the longitudinal nature of the data. Alternative approaches to modeling outcomes in Study 1, for example, using Structural Equations Modeling (SEM), could model trajectories over time, and use trajectory slopes as predictors. For example, such an approach would allow me to ask whether people whose motivation trajectories were similar (e.g., as derived from latent class analysis of trajectories, characterized by both intercept and slope) were more or less subjectively successful. There are three main reasons I did not model time or trajectories explicitly.

The first reason I did not model time or trajectories is that the data structure here is not easily accommodated by SEM in ways that could preserve within-person data. The crucial issue here is that in these data, observations (i.e., goals) are nested within people, but variably so. Further, this variation is meaningful and not due to missing data. SEM can accommodate non-independent observations, and is ideal for doing so in many contexts, such as in assessing measurement or in modeling intensively sampled phenomena using Ecological Momentary Assessment (EMA) However, in SEM, variation in the number of observations per person must be modeled as equal. When observations within a person are missing in a dataset, such as when people have failed to respond to an EMA prompt, treating “missing” observations as missing is valid. However, when observations within a person are missing because they do not exist, as is the case for people who had fewer than five resolutions, treating “missing” resolution observations as missing is not valid. An alternative approach to modeling each resolution within each person would be to compute person-level scores (i.e., to average observations about each resolution within each person). Although this is conceptually similar to the modeling process of multilevel regression, and although it resolves the issue of variable numbers of resolutions per person, it introduces more issues, such that estimate standard errors would vary as a function of the number of resolutions people had and it would preclude analysis of within-person dynamics.

Although modeling trajectories in SEM was not feasible, there were other ways I could model time (e.g., as a third level in the multilevel models). A second reason I did not model time or trajectories explicitly is that doing so would have increased the potential bias due to missing data, whether in SEM or in a multilevel regression. As explained in detail in the next section, my modeling approach allowed missing data to have a minimal potential impact on results, because my focus was almost exclusively on variables measured at the beginning of the year. Attrition increased for each wave of the survey, and analyses that relied heavily on data from later time points would increase my reliance my (imperfect) approaches for handling missing data and the assumptions that support the use of those approaches.

A third reason I did not model time or trajectories explicitly is that it would entail a large number of arbitrary, but potentially consequential, analytic decisions. The focal variables here were not all measured at every time point, which would require that I devise an analytic approach anew for each goal outcome. For example, some variables were binary, and others were continuous. Binary and continuous variables require different kinds of trajectory analyses, and further, trajectory analyses do not make sense for all goal-varying properties (e.g., domain). Explicitly modeling time or trajectories would not only increase analytic flexibility, but it would also make presenting, comparing, and interpreting analyses more difficult.

2.2.4.3 Missing data

Study 1 was a longitudinal study administered by survey four times over the course of a year. People were compensated for completion of each survey regardless of their completion of others. To reduce attrition, participants received bonuses based on the number of surveys they completed, were given windows of about ten days to complete the surveys, and received reminder emails if they had not completed an active survey. Despite these efforts, there was attrition throughout the year. The missing data resulting from attrition limits my ability to interpret patterns involving data from later surveys at face value (particularly descriptive statistics). There are many approaches for handling missing data among independent variables, although no approach is without assumptions and caveats.

The appropriate missing data approach for independent variables in a given analytic situation depends on why data are missing and how problematic missingness is (Rubin, 1976). Missingness is least problematic when it is completely random. When missing values are random, they can be handled with listwise deletion (i.e., removal of observations with missing data) without biasing estimates of parameters. For example, if only a subset of participants were invited to participate in follow-up surveys on the basis of random selection, and everyone invited to complete follow-up surveys did, listwise deletion would increase variance but would not bias estimates. Missingness is most problematic (and cannot be statistically remedied) when it is caused by the missing values themselves. For example, if participants in this study wanted to avoid reporting failure and selectively skipped questions about their progress on failing resolutions but answered questions about their progress on successful resolutions, then missingness on the progress variable would be caused by the missing values. In other situations, missingness is problematic, but not catastrophically so. When missingness can be accounted for by other information in the dataset (assuming missingness is not caused by the missing values themselves), simulation studies suggest that bias in parameter estimates and standard errors introduced by missingness can be statistically remedied, for example using maximum likelihood estimation or multiple imputation (Enders, 2017; Schafer & Graham, 2002). These approaches allow people without complete data to be included in analyses.

About 60% (n = 250) of participants completed the final survey and therefore reported on the outcome of their goal pursuit. The most common pattern of survey completion was completing all four waves (n = 199), followed by completing only the first survey (n = 69). Every possible pattern of survey completion was represented in the data. About as many people completed two surveys total (n = 77) as completed three surveys total (n = 70).

Missingness at the level of survey could be predicted by other variables in the dataset. The number of surveys people completed was positively associated with all individual difference measures of and closely related to self-regulatory skill: Conscientiousness (r = 0.197, $t(405)$ = 4.038, p = 0.001); Trait Self-Control (r =

0.145, $t(414) = 2.973$, $p = 0.009$); Grit ($r = 0.203$, $t(414) = 4.208$, $p = <.001$); Impulsivity ($r = 0.173$, $t(414) = 3.563$, $p = 0.002$). One exception was the measure of goal disengagement capacity, which is designed to measure a self-regulatory skill but did not have convergent validity in this sample (i.e., it was not correlated with any other individual difference measures). The association between goal disengagement capacity and number of surveys completed was not distinguishable from zero ($r = -0.011$, $t(414) = -0.227$, $p = 0.82$). In addition, age was modestly positively correlated with the number of surveys completed ($r = 0.226$, $t(414) = 4.705$, $p = <.001$). The number of surveys completed by people did not differ as a function of education level whether using the levels as reported (some of which were sparse; $F(6,408) = 1.613$, $p = 0.142$) or more coarse levels ($F(2,412) = 2.018$, $p = 0.134$).

Missingness should not bias the descriptive analyses I present if not accounted for, largely because it is very minimal among the independent variables of interest, which were mostly collected at the first wave of data collection. Thus, for most descriptive analyses, missing data need not be accounted for among independent variables. For analyses that do (i.e., inferential analyses and descriptive analyses that use predictors from later in the year), maximum likelihood estimation is warranted and can be implemented in the context of (frequentist) mixed effect regressions. However, missingness handled with listwise deletion does increase variance of parameter estimates and statistical power due to reduced sample size. Note that because missingness is at the level of the survey and person, not only is it unlikely to result from pointed omission of specific information, but it also primarily threatens inferences about between-subjects effects rather than with respect to variation among people's goals. Thus, missing data threatens to bias parameters only to the extent that (mid-year) factor effects vary as a function of individual differences, which can be reasonably estimated in the models (albeit imperfectly, given the missingness). For analyses that do have missing values among predictors, I use maximum likelihood estimation in frequentist mixed effect regressions.

2.3 Results

2.3.1 Research question 1: Characterizing resolution content and properties

Most people reported three or fewer resolutions (n_{people} = 317), out of a maximum possible of five. The modal number of reported resolutions was two (n_{people} = 113). Most resolutions (66.3%) were totally new and had never been set before as resolutions (*n* = 725). The number of resolutions in this study was calculated by counting how many resolutions people reported, rather than by directly asking people to report the number of resolutions they had set.

2.3.1.1 Content of resolutions

2.3.1.1.1 Word counts

People described their resolutions in open-ended texts, which were assessed with automated word counts. The most three most common word stems all related to physical health. Among the top 30 word stems were those that related to finances ("money," *n* =63; "save," *n* = 57; "spend," *n* = 35), family ("family," *n* = 31), and smoking ("smoke," *n* = 40). An estimated 6% (*n* = 69) of resolutions pertained to substance use (i.e., included word stems related to addiction, tobacco use, or alcohol use).

People also provided open-ended descriptions of their reasons for pursuing resolutions. The most common word stems in responses suggested that people's reasons for setting resolutions related to their subjective experiences ("feel"), physical health ("weight" and "health"), and finances ("money"). People's responses also suggested abstract thinking ("life" and "time").

Table 1: Top 10 Word Stems in New Year's Resolutions and Reasons in Study 1

	Resolutions		**Reasons**	
Rank	Word	*n*	Word	*n*
1	Lose	134	Feel	118
2	Weight	115	Weight	115
3	Exercise	74	Health	109
4	Time	72	Money	99
5	Eat	63	Time	95
6	Money	63	Life	84
7	Save	57	Set	76
8	Pound	49	Lose	66
9	Start	43	Resolution	66
10	Week	41	Healthier	57

Note. Stop words were removed and all words were stemmed. Word stems were edited for readability (e.g., "exercis" as "exercise").

2.3.1.1.2 Domains

People who completed at least one follow-up survey were asked to report the non-exclusive life domains their resolution related to: physical health, mental health, money, career, social, family, society, spiritual education and other. The most common domains were Physical (n = 477) and Mental (n = 420) and the least was Education (n = 64). As shown in Table 2, considering combinations of selected domains, the most common pattern was physical health as the only selected domain, followed by physical health and mental health, and then money as the only selected domain. Of the 24 resolutions that were associated with a domain not listed, the words that appeared more than once across people's open-ended domain descriptions were: "personal" (n = 3), "development" (n = 2), "enjoyment" (n = 2), "hobby" (n = 2), and "home" (n = 2).

In subsequent analyses, of the nomothetic domain categories that people reported, I assessed affects associated with only physical and mental domains, as they were the two most common and were independent from one another, but each roughly split the sample.

Table 2: Frequencies of Patterns in Life Domains Associated with at Least 20 New Year's Resolutions in Study 1

Rank	Physical	Mental	Money	Career	Social	*n*
1	✓					171
2	✓	✓				107
3			✓			69
4		✓				28
5			✓	✓		27

Note. Among patterns of life domains associated with at least 25 resolutions, none related to the five other domain options: Family, Society, Spiritual, Education and Other.

2.3.1.2 Research question 1a: Concreteness

The clear majority – over 90% – of resolutions were categorized by human coders as concrete (e.g., "I resolve to drink less beer"; n = 1007) rather than abstract (e.g., "to love my wife more"; n =81).

2.3.1.2 Research question 1b: Specificity

Coding specificity was challenging given how short most resolution texts were. Fewer than half of resolutions (40%) were coded as having a specific end point (e.g., "to get a new job"; n = 435). More often, resolutions were not specific (60%). They either had a relative objective, and thus a non-specific end-point (e.g., “losing weight"; *n* = 507) or a subjective objective and thus a vague end-point (e.g., “to become more emotionally free and independent from men mostly, and others"; *n* = 148).

2.3.1.3 Research question 1c: Approach, avoidance, maintenance

Coding approach, avoidance, and maintenance based on the short texts of resolutions was often challenging. For example, the goal to lose weight was coded as an approach goal because in setting a goal to lose weight, people are approaching a desired state (being a lower weight) rather than avoiding an undesired state (e.g., as would be the case if the goal was to avoid gaining weight).

Most resolutions (86%) involved approaching a desired state (e.g., “Find love.”; n = 936) rather than avoiding an undesired state (e.g., “I have resolved to quit smoking completely by the end of February.”; *n* = 110). While many resolutions were implicitly maintenance goals in that the desired (or undesired) behavior would continue (or continue to stop) indefinitely, a small handful of goals were explicitly maintenance goals in that the goal objective related to a current desired or undesired behavior pursuer wanted to maintain (e.g., Keep going on the transfer to 4 years college" *n* = 8). In addition, some resolutions described a combination of approach and avoidance (e.g., “Eat a Healthier Diet - remove gluten and sugar from my diet”; *n* = 35).

2.3.2 Research question 2: Characterizing motivation

Consistent with high reliability at each time point, as reported in the measures section, correlations among commitment, effort, and confidence were high in January (see Table 1) and remained high in April (T2; r range = 0.67 to 0.75, all $p < .001$) and July (T3; r range = 0.62 to 0.80, all $p < .01$).

2.3.2.1 Motivation at the beginning of the year

In January, average Motivation was high ($M = 4.19$, $SD = 0.82$, $N = 1094$). For the plurality of resolutions, most people were "very" committed ($n = 682$; 62%), dedicating "a lot" of effort ($n = 499$; 46%), and were "completely" confident that they would achieve their resolution within the calendar year ($n= 506$; 46%).

Table 3: Commitment, Effort, and Confidence Means, Standard Deviations, And Correlations at the Beginning of the Year (T1) in Study 1

Variable	*M*	*SD*	1	2
1. Commitment	4.45	0.81		
2. Effort	4.01	1.13	.63**	
3. Confidence	4.10	1.03	.54**	.45**

Note. Commitment, effort, and confidence were one-item measures. People reported how committed they were to their resolution on a Likert scale from 1 (*not at all*) to 5 (*very*), and how confident they were that they would achieve their resolution by the last day of the year on a Likert scale from 1 (*not at all*) to 5 (*very*). *M* and *SD* are used to represent mean and standard deviation, respectively. * indicates $p < .05$. ** indicates $p < .01$.

2.3.2.2 Average change in motivation throughout the year

Compared to the beginning of the year, Motivation for resolutions was lower in both April (M = 3.79, SD = 1.15, N = 809) and July (M = 3.58, SD = 1.23, N = 656). On average, Motivation lowered over time, as shown in Figure 1.

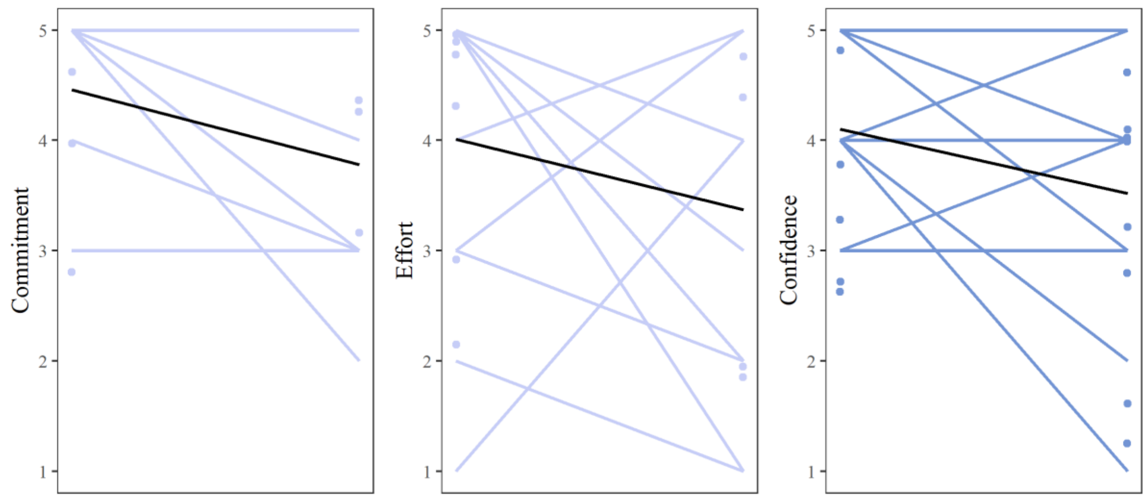

Figure 1: Motivation Change from January (T1) to July (T3) in Study 1. Changes from January to July in commitment, effort, and confidence for 15 randomly selected resolutions. In each panel, the black line is fit to the average intercept and slope of the effect of time.

The ICC of the Motivation composite in July (T3) was 0.12, indicating that approximately 12% of variance in Motivation in July was accounted for by clustering of resolutions within people. As shown in Table 4, the positive effect of early Motivation on mid-year Motivation operated both within and between people, over and above Trait Self-Control. Further, an ANOVA comparing models with and without random slopes suggests that slope variance was marginally significant ($\tau_{11} = 0.21$, $p = 0.041$) suggesting that people differed in the extent to which their early Motivation predicted their later Motivation. Unsurprisingly given skewed responses at the beginning of the year, intercepts were negatively correlated with slopes.

Table 4: The Effect of April Motivation (T1) on July Motivation (T3) in Study 1

Predictors	*Estimates*	*CI*
(Intercept)	3.53 ***	3.44,3.62
Motivation (goal)	0.29 **	0.11,0.48
Motivation (person)	0.51 ***	0.34,0.69
Trait Self-Control	0.29 ***	0.16,0.41
Random Effects		
σ^2	1.16	
τ_{00}	0.08	
τ_{11}	0.21	
ρ_{01}	-0.04	
N	253	
Observations	656	
Marginal R^2 / Conditional R^2	0.134 / 0.227	

Note. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random intercepts and slopes. *$p<0.05$ ** $p<0.01$ *** $p<0.001$*

2.3.2.2 Motivation trajectories

Average Motivation dropped throughout the year, but, as signaled by the marginally significant slope variance in the multilevel model and shown in Figure 2, trajectories of individual resolutions over time did not uniformly drop from January to April to July. Many trajectories were nonmonotonic (n = 775) or stable (n = 123) rather than monotonic and steadily decreasing (n = 248). Autocorrelations across the year were modest (Table 5).

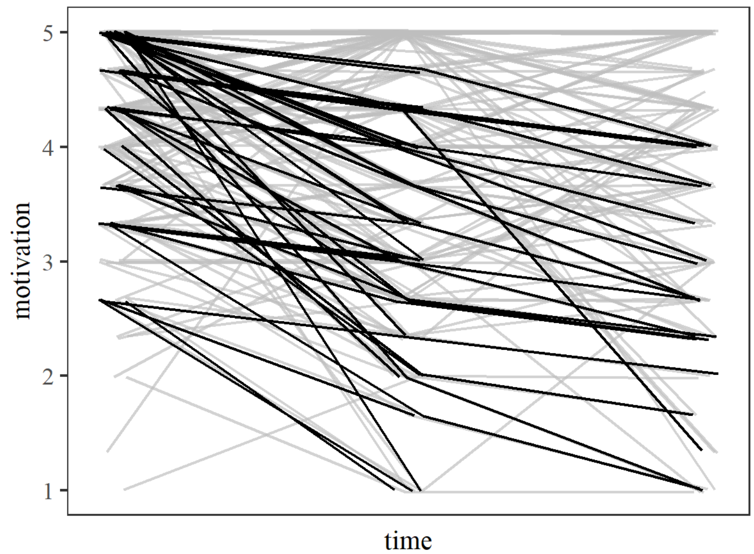

Figure 2: Motivation Trajectories of 100 Randomly Selected Resolutions from January (T1) to April (T2) to July (T3) in Study 1 with Decreasing Trajectories Highlighted

Table 5: Commitment Means, Standard Deviations, and Correlations with Confidence Intervals in January (T1), April (T2), and July (T3) in Study 1

Variable	*M*	*SD*	1	2
1. Motivation T1	4.45	0.81		
2. Motivation T2	4.02	1.19	.34**	
3. Motivation T3	3.82	1.28	.25**	.45**

Note. Commitment was a one-item measure. People reported how committed they were to their resolution on a Likert scale from 1 (*not at all*) to 5 (*very*). *M* and *SD* are used to represent mean and standard deviation, respectively. Values in square brackets indicate the 95% confidence interval. * indicates $p < .05$. ** indicates $p < .01$.

2.3.3 Research question 3: Characterizing pursuit

People's pursuit behaviors were consistent with the domains they indicated. As shown in Table 6, many of the most common word stems and bigrams clearly related to physical health goals and money goals.

Table 6: Top 10 Word Stems and Stemmed Bigrams in Behaviors Associated with Pursuit in April (T2) in Study 1

Rank	Word Stems	*n*	Bigrams	*n*
1	Eat	165	Eat Healthy	25
2	Time	153	Save Money	24
3	Exercise	103	Spend Time	11
4	Daily	97	Healthy Food	9
5	Money	96	Physical Activity	8
6	Food	75	Credit Card	7
7	Week	55	Junk Food	7
8	Spend	54	Positive Attitude	6
9	Walk	54	Savings Account	6
10	Drink	53	Time Management	6

Note. Stop words were removed and all words were stemmed. Word stems were edited for readability (e.g., "healthi" and "healthier" as "healthy").

A handful of resolutions were identified by word counts as having an explicit frequency (i.e., they contained words like "every day"). Relatively few goals (n = 82) were intended to be pursued at a particular frequency: either daily (e.g., "I resolved to spend at least 15 minutes per day writing jokes and posting at least one each day on Twitter"; n = 45) or weekly (e.g., "My goal is to work out and get in shape. I have a goal of going to the gym three days a week for an hour a trip"; n = 37).

2.3.3.1 Research question 3a: Social commitment

In April, Social Commitment as measured with the item "Others know that I have made this resolution" was nearly multi-modal as shown in Figure 3. The true modal response was 1 (*not at all true*) and the second mode was at the far end of the scale at 5 (*completely true*). Social in July was similarly multi-modal, and was correlated with early Social Commitment within people's first resolutions (r = 0.50, p < .001).

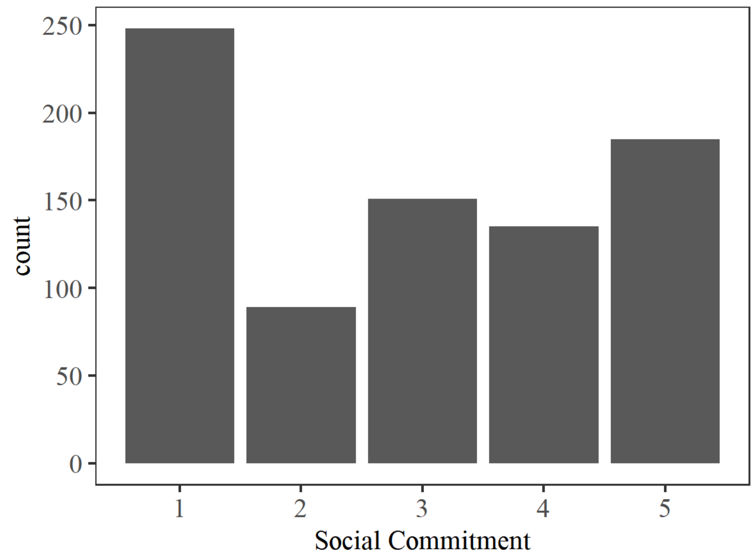

Figure 3: Histogram of Social Commitment in April (T2) in Study 1

2.3.3.2 Habit formation

A key characteristic of habit is daily or near-daily frequency in a stable location. People reported the frequency and context stability of their resolution-related behaviors, and the product of these measures is a standard measure of Habit Formation.

As shown in Table 4, many resolutions were at or above the midpoint in goal pursuit frequency (M =3.34, SD = 1.24) on a scale from 1 (*never or almost never*) to 5 (*everyday*). As shown in Table 5, many resolutions were at or above the midpoint in goal pursuit location stability (M =3.79, SD = 1.16) on a scale from 1 (*never*) to 5 (*always*). Frequency was modestly correlated with stability among people's first resolutions (r = 0.35, p < .001).

Habit Formation was moderate in April (T2; M =13.62 of a maximum possible 25, SD = 6.11). Responses in July showed similar patterns; within people's first resolutions, frequency and stability in July were autocorrelated with April responses (r =

0.30 and $r = 0.34$, respectively, both $p < .001$) and were similarly correlated with one another in July as in April ($r = 0.34$, $p < .001$). Habit Formation in July was correlated with Habit Formation in April ($r = 0.35$, $p < .001$).

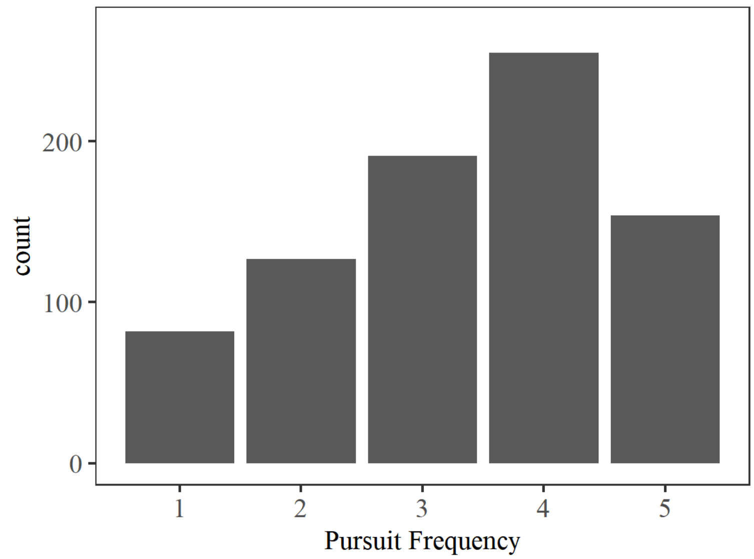

Figure 4: Histogram of Pursuit Frequency in April (T2) in Study 1

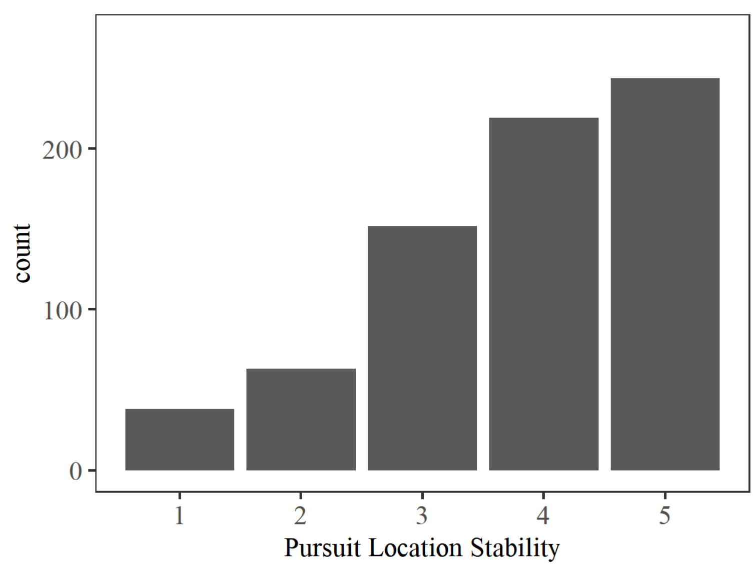

Figure 5: Histogram of Pursuit Location Stability in April (T2) in Study 1

2.3.4 Research question 4: Characterizing success

People reported the status of their resolutions throughout the year. In April, most people were actively pursuing their resolutions (n = 585) or their resolutions were "on hold" (n = 194). Very few resolutions had not been started (n = 12) or had been disengaged from (n = 16).

In January, of those who completed the final survey, the plurality of resolutions were achieved, including those that were achieved and being maintained (n = 250). Many other resolutions were still actively being pursued (n = 204). Of the rest, most were on hold (not being pursued, with plans to pursue later; n = 140), disengaged from (n = 35) or never started, with plans to start later (n = 8).

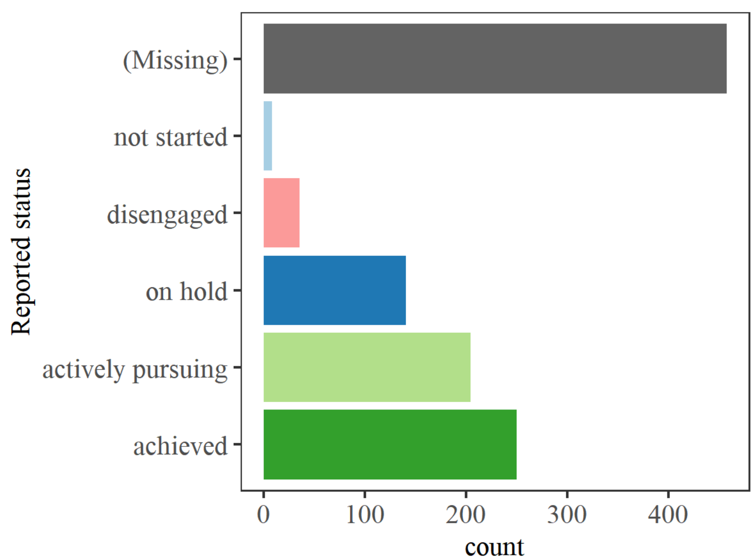

Figure 6: Histogram of Final Resolution Status (T4) in Study 1

At the end of the year, people also provided continuous ratings of their Subjective Success and their objective achievement. On average, resolutions were achieved to a moderate degree, but there was substantial variability in self-reported achievement (M =

2.99, *SD* = 1.43). The distribution of achievement was nearly uniform. Subjective Success ratings were similar in point estimate and variability (*M* = 2.98, *SD* = 1.43), and had a similarly flat distribution.

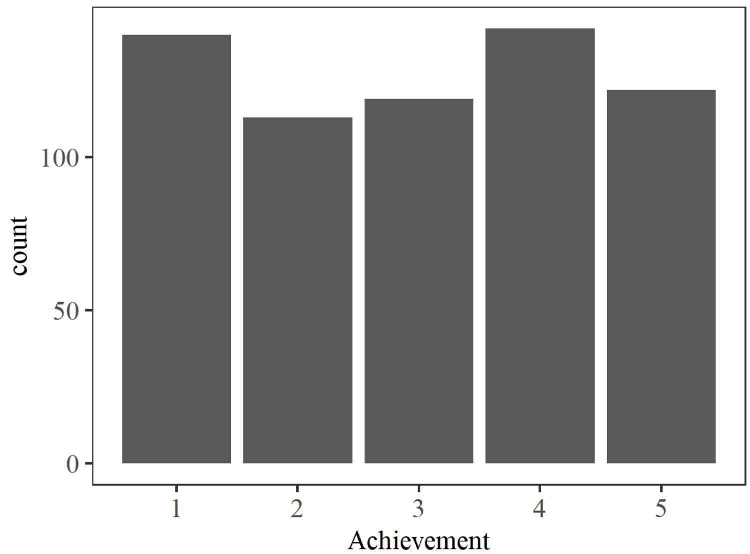

Figure 7: Histogram of Achievement (T4) in Study 1

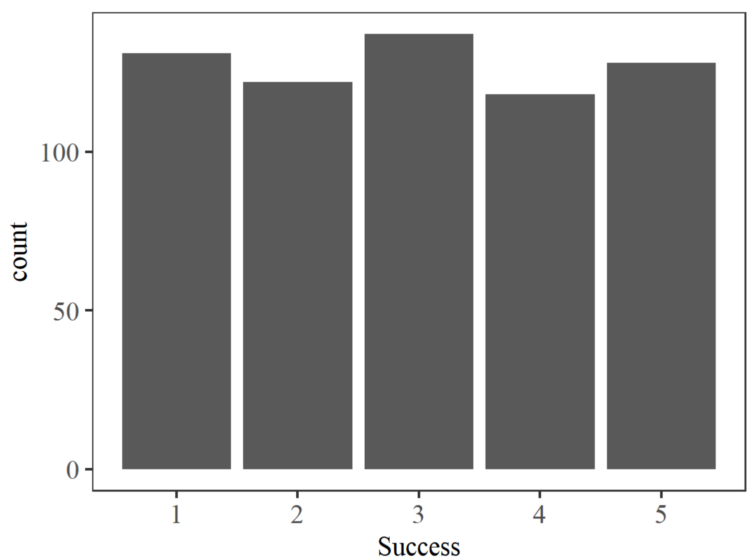

Figure 8: Histogram of Subjective Success (T4) in Study 1

These various measures of success were similar to one another. Achievement and Subjective Success were highly correlated among people's first resolutions ($r = 0.83$, $t(633) = 23.173$, $p < .001$). Further, people's ratings of their achievement and Subjective Success corresponded to people's reported status.

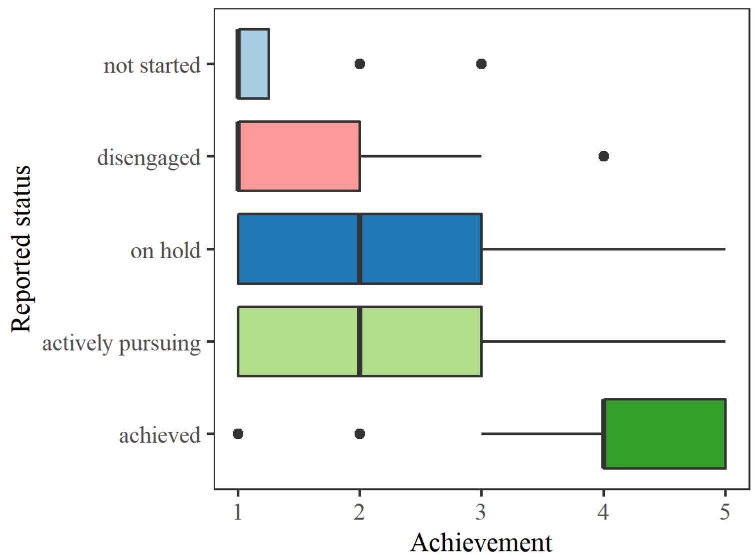

Figure 9: Achievement (T4) by Final Resolution Status (T4) in Study 1

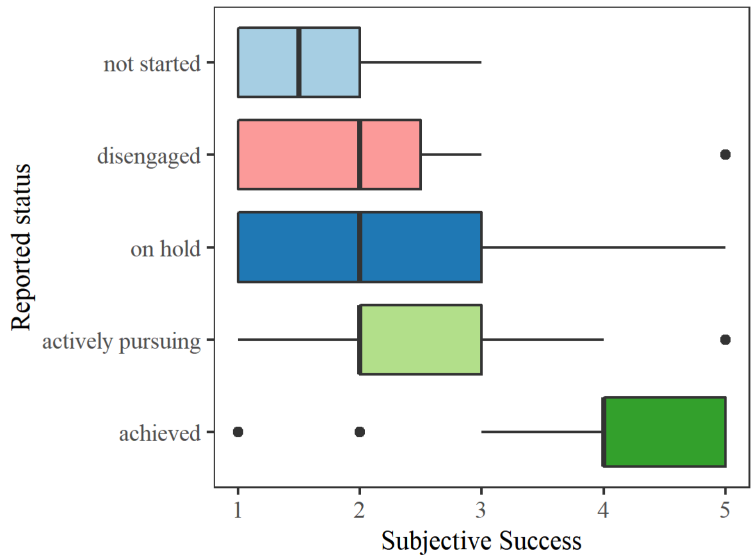

Figure 10: Subjective Success (T4) by Final Resolution Status (T4) in Study 1

2.3.5 Research question 5: Characterizing disengagement

Few people deliberately disengaged from their goals (*n* = 63, 6%). Further, disengagement was unstable over time despite the fact that it was operationalized as having stopped pursuit, with no plans of ever continuing. Of people the 16 resolutions that were disengaged in April (T2), only 5 were still disengaged in July (T3; 3 were missing), and only 3 were still disengaged in January (T4; 3 were missing). As shown in Figure 11, most people did not think about quitting their resolution (M = 2.23; SD = 1.47, N = 636).

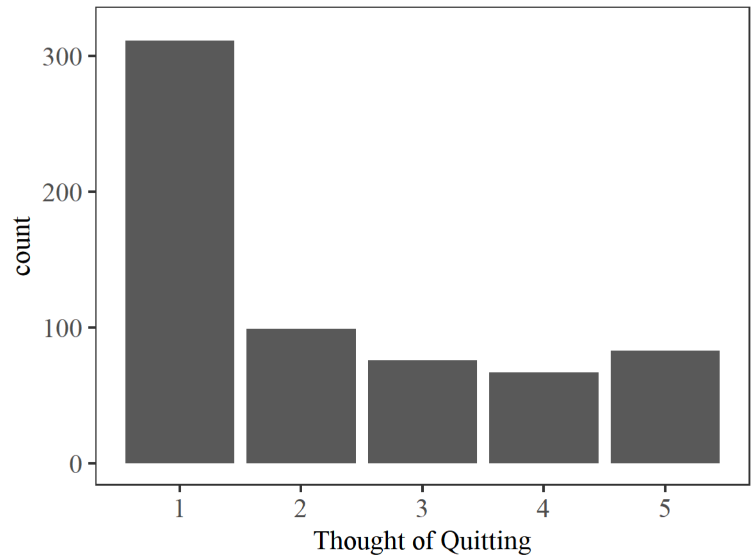

Figure 11: Histogram of Thought of Quitting in Study 1

2.3.6 Research question 6: Characterizing covariation

Pearson *r* correlations based on people's first resolutions are shown in Table 7. Pearson *r* correlations based on the entire dataset are shown in Appendix A. All follow-up tests conducted to investigate covariance were conducted on a reduced sample of people's first resolutions only.

Table 7: Means, Standard Deviations, N, and Correlations of Goal-Varying Properties Derived from People's First New Year's Resolution in Study 1

Variable	*M*	*SD*	*N*	1	2	3	4	5	6	7	8
1. Physical Domain	1.69	0.46	331								
2. Mental Domain	1.43	0.50	331	-.03							
3. Specific-Relative	0.41	0.49	413	.11*	-.02						
4. Specific-Vague	0.11	0.31	413	-.02	.10	-.29**					
5. Concrete-Abstract	1.05	0.22	411	-.10	.17**	-.17**	.63**				
6. Avoid-Approach	0.87	0.33	394	-.16**	.12*	.20**	.06	-.06			
7. Motivation	4.30	0.72	415	-.15**	-.01	-.06	-.06	-.03	.12*		
8. Habit Formation	14.01	5.83	284	.01	-.04	.02	-.01	-.01	.02	.25**	
9. Social Commitment	3.05	1.46	311	.06	-.14*	.05	-.06	-.09	.03	.23**	.17**

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. *N* is used to indicate the available sample size for each variable. Variables derived from the survey administered in January (T1) were Specificity, Concrete-Abstract, Avoid-Approach, and Motivation. Dummy coded variables list the level coded as zero first and the level coded as one second. Specific-Relative and Specific-Vague are dummy coded Specificity variables. * indicates $p < .05$. ** indicates $p < .01$.

As shown in Table 7, Social Commitment was associated with more Motivation and more habitual goal pursuit. In addition, more Motivation at the beginning of the year was associated with more habitual pursuit months later.

As shown in Table 8, for most variables, ICCs were relatively large, ranging from .20 to .40. For Avoid-Approach and Physical Domain, ICC was not calculated directly but the numerator of the ICC, the estimated random intercept variance, was zero.

Table 8: ICCs and Number of People and Observations for Goal-Varying Properties in Study 1

Variable	*ICC*	*N* people	*N* resolutions
Physical Domain	0*	331	854
Mental Domain	0.30	331	854
Specificity	0.20	413	1090
Concrete-Abstract	0.33	411	1088
Avoid-Approach	0*	394	1046
Motivation	0.24	415	1094
Habit Formation	0.24	284	716
Social Commitment	0.40	311	808

Note. *ICC* represent the intraclass correlation estimated from a model with a random intercept for people. * indicates that the estimated ICC was zero because the estimated random intercept variance was zero.

Categorical goal properties covaried. Most striking was the overlap between concreteness and specificity shown in Table 8. Abstract goals were less specific than concrete goals, $\chi^2(2) = 165.49$, $p < .001$. Both of these codes captured how objectively measurable an outcome was and were somewhat redundant such that all abstract goals were vague and non-specific and none were specific, even in the entire dataset.

Table 9: Frequencies of Concrete-Abstract and Specificity Among First Resolutions with Examples in Study 1

	Specific	Non-Specific	Vague
Concrete	197 “watch every Meryl Streep movie ever made”	169 “lower blood glucose level”	24 “get super fit"
Abstract	0	1 " reduce my level of stress"	20 “figure out my life"

Goals that related to physical health had a marginally significant point biserial correlation with one of the specificity dummy coded variables, but a chi-square examining all three levels of specificity simultaneously was not significant, $\chi^2(2) = 4.23$, $p = .121$.

Physical and non-physical goals differed in their avoidance and approach focus, $\chi^2(1) = 7.81$, $p = .005$. As shown in Table 10, physical goals were more often avoid-focused. Or, equivalently, relatively more avoidance-focused goals were related to physical health than approach-focused goals.

Table 10: Frequencies of Avoid-Approach and Physical Domain Among First Resolutions with Examples in Study 1

	Physical	Not Physical
Avoidance	38 “Have fewer than 10 drinks per week”	5 “…ignore people who disagree with my politics on social media….”
Approach	182 “I want to exercise more”	88 “save 10% of my income”

Goals that were related to mental health differed in approach- and avoidance-focus compared to those that weren't related to mental health, $\chi^2(1) = 4.16$, $p = .041$. As shown in Table 11, mental goals were more often approach-focused than were non-mental goals. Equivalently, avoidance goals were relatively less often related to mental health than approach goals.

Goals that were related to mental health differed in concreteness from those that weren't related to mental health, $\chi^2(1) = 7.87$, $p = .005$. As shown in Table 12, mental goals were more often abstract than were non-mental goals. Equivalently, abstract goals were relatively more often related to mental health than concrete goals.

Table 11: Frequencies of Avoid-Approach and Mental Domain Among First Resolutions with Examples in Study 1

	Mental	Not Mental
Avoid	12 "stop yelling at my kids"	31 "reduce needless spending"
Approach	120 "I want to get back into work on my art projects every day"	150 "go to the doctor for a physical exam"

Table 12: Frequencies of Concrete-Abstract and Mental Domain Among First Resolutions with Examples in Study 1

	Mental	Not Mental
Concrete	128 “do yoga 30 minutes daily”	183 “to pay down at least 50% of our household credit card debt by the end of the year”
Abstract	14 “I want to discover what it is I want to do with my life”	4 " I will be a better husband"

Approach-focused goals differed in specificity from avoidance-focused goals, $\chi^2(2) = 20.87$, $p <.001$. As shown in Table 13, compared to approach-focused goals, relatively more avoidance-focused goals were specific. Said differently, approach-focused goals were more often non-specific or vague than avoidance-focused goals.

Table 13: Frequencies of Avoid-Approach and Specificity Among First Resolutions with Examples in Study 1

	Specific	Non-Specific	Vague
Avoid	39 “quit smoking”	8 “drink less alcohol”	3 “I resolved to stop my obsessive thinking"
Approach	150 “fix car”	156 "to eat a more nutritious diet"	38 “I plan on starting a career"

Some categorical goal properties also covaried with Motivation, Habit Formation, and Social Commitment, as shown in Table 7. Physical goals were associated with less Motivation at the beginning of the year compared to non-physical goals. Approach goals were associated with more Motivation at the beginning of the year. Mental goals were associated with less Social Commitment compared to non-mental goals.

2.3.7 Research question 7: Associations between skill in self-regulation and goal-varying factors

As shown in Table 14, among people's first resolutions, Trait Self-Control was positively associated with goal specificity, initial Motivation, and Habit Formation. Trait Self-Control differed among people with specific (M = 3.58, SD = 0.77, n = 198), non-specific (M = 3.58, SD = 0.77, n = 171), and vague (M = 3.25, SD = 0.83, n = 44) resolutions, $F(2,410) = 4.09$, $p = .017$. Post hoc pairwise comparisons with Tukey corrections revealed that people with specific first resolutions had significantly higher average Trait Self-Control than those with vague resolutions, $p = 0.027$. Trait Self-Control was associated with more Motivation at the beginning of the year and more Social Commitment mid-year among people's first resolutions.

Table 14: Correlations of Each Goal-Varying Property with Trait Self-Control Derived from People's First Resolutions in Study 1

Variable	*r*	CI
Physical Domain	-.05	-.15, .06
Mental Domain	-.06	-.16, .05
Specific-Relative	-.06	-.15, .04
Specific-Vague	-.10*	-.20, -.01
Concrete-Abstract	-.04	-.14, .05
Avoid-Approach	.03	-.07, .13
Motivation	.30**	.21, .38
Habit Formation	.12*	.00, .23
Social Commitment	.10	-.01, .21

Note. Trait Self-Control M = 3.48, SD = 0.76, N = 415. * indicates $p < .05$. ** indicates $p < .01$.

2.3.8 Research question 8: Explaining variance in success

To provide context for regression analyses, correlations first were estimated between each predictor variable and Subjective Success at the end of the year and the relationship between Trait Self-Control and success was estimated in a multilevel regression. Then, sets of multilevel regressions were conducted to predict Subjective Success from each goal-varying property accounting for Trait Self-Control. The ICC of success was 0.09, suggesting that about 9% of variance in success is attributable to person-level variance.

Table 15: Correlations of Each Goal-Varying Property with Subjective Success Derived from People's First Resolutions in Study 1

Variable	*r*	CI
Physical Domain	-.03	-.15, .10
Mental Domain	.04	-.09, .17
Specific-Relative	.06	-.07, .18
Specific-Vague	-.03	-.15, .10
Concrete-Abstract	-.02	-.15, .10
Avoid-Approach	.04	-.09, .17
Motivation	.14*	.02, .26
Habit Formation	.20**	.07, .33
Social Commitment	.07	-.06, .20

Note. Success M = 2.92, SD = 1.37, N = 250. * indicates $p < .05$. ** indicates $p < .01$.

As shown in Table 16, skill in self-regulation was associated with success at the end of the year. People with average Trait Self-Control were estimated to be moderately successful. For each one-unit increase in Trait Self-Control, the model estimated a corresponding .26-unit increase in Subjective Success.

Table 16: Multilevel Model Regressing Trait Self-Control on Success in Study 1

Predictors	*Estimates*	*CI*
(Intercept)	2.97 ***	2.85,3.08
Trait Self-Control	0.26 ***	0.12,0.41
Random Effects		
σ^2	1.83	
τ_{00}	0.13	
N	250	
Observations	636	
Marginal R^2 / Conditional R^2	0.022 / 0.086	

* *p<0.05* ** *p<0.01* *** *p<0.001*

Note. σ^2 represents residual variance. τ_{00} represents random intercept variance.

2.3.6.1 Goal domain

The two most common domains, physical health and mental health, were each separately regressed on Subjective Success with random intercepts and slopes. For both physical domain and mental domain, the maximal model with random slopes failed to converge, so random slopes were not estimated.

As shown in Table 17, there was insufficient evidence that whether a goal related to physical health or not explained variance in success. The fixed effects of this model explained an estimated 2.6% of variance in success and the fixed and random effects explained an estimated 10.5% of variance.

Table 17: Multilevel Model Regressing Physical Domain on Subjective Success in Study 1

Predictors	*Estimates*	*CI*
(Intercept)	2.93 ***	2.76,3.11
Physical	0.06	-0.17,0.28
Trait Self-Control	0.29 ***	0.14,0.44
Random Effects		
σ^2	1.80	
τ_{00}	0.16	
N	235	
Observations	604	
Marginal R^2 / Conditional R^2	0.026 / 0.105	

Note. CI represents confidence interval. Physical health domain was coded as one if present and zero if absent. σ^2 represents residual variance. τ_{00} represents random intercept variance. *p <0.05 ** p <0.01 ***p <0.001

As shown in Table 18, there was insufficient evidence that whether a goal was related to mental health or not explained variance in success. The estimated percent of variance explained by fixed and all effects were similar in this model as in the model with physical health.

Table 18: Multilevel Model Regressing Mental Domain on Subjective Success in Study 1

Predictors	*Estimates*	*CI*
(Intercept)	2.96 ***	2.79,3.13
Mental	0.01	-0.22,0.24
Trait Self-Control	0.29 ***	0.13,0.44
Random Effects		
σ^2	1.80	
τ_{00}	0.16	
N	235	
Observations	604	
Marginal R^2 / Conditional R^2	0.026 / 0.103	

Note. CI represents confidence interval. Mental health domain was coded as one if present and zero if absent. σ^2 represents residual variance. τ_{00} represents random intercept variance. **p* <0.05 ** *p* <0.01 ****p* <0.001

2.3.6.1 Concrete-abstract

Goal concreteness was regressed on Subjective Success. The maximal model with random slopes failed to converge, so random slopes were not estimated. As shown in Table 19, there was insufficient evidence that whether a goal was concrete or abstract explained variance in success longitudinally. The fixed effects in the model explained an estimated 2.3% of the variance in success and the fixed and random effects explained an estimated 9.5% of variance in success.

Table 19: Multilevel Model Regressing Concrete-Abstract on Subjective Success in Study 1

Predictors	*Estimates*	*CI (95%)*
Intercept	2.97 ***	2.85,3.09
Concrete-Abstract	-0.17	-0.62,0.28
Trait Self-Control	0.26 ***	0.11,0.41
Random Effects		
σ^2	1.81	
τ_{00}	0.14	
N	248	
Observations	631	
Marginal R^2 / Conditional R^2	0.023 / 0.095	

Note. CI represents confidence interval. Concrete was coded as zero and abstract was coded as one. σ^2 represents residual variance. τ_{00} represents random intercept variance. *p <0.05 ** p <0.01 ***p <0.001

2.3.6.1 Specificity

The effect of goal specificity on success was estimated by regressing two dummy variables comparing specific goals to non-specific goals and specific goals to vague goals on Subjective Success. The maximal model with random slopes failed to converge, so random slopes were not estimated. A test comparing models with and without the specificity dummy variables suggested that model fit was not improved by the inclusion of specificity information $\chi^2(2) = 4.45$, $p = 0.11$. As shown in Table 20, the fixed effects of the model with specificity variables explained less than 3% of variance in success and the fixed and random effects explained an estimated 9.6% of the variance in success.

Table 20: Multilevel Model Regressing Specificity on Subjective Success in Study 1

Predictors	*Estimates*	*CI (95%)*
Intercept	2.83 ***	2.65,3.00
Specific-Non-Specific	0.22	-0.02,0.45
Specific-Vague	0.29	-0.06,0.65
Trait Self-Control	0.27 ***	0.13,0.42
Random Effects		
σ^2	1.82	
τ_{00}	0.13	
N	248	
Observations	632	
Marginal R^2 / Conditional R^2	0.029 / 0.096	

Note. CI represents confidence interval. Specificity was coded as zero for both dummy-coded variables. σ^2 represents residual variance. τ_{00} represents random intercept variance. *p <0.05 ** p <0.01 ***p <0.001

2.3.6.1 Avoid-approach

Very few goals were maintenance goals or specified a combination of approach, avoidance, and maintenance, so I compared just approach-focused goals to avoidance-focused goals. As shown in Table 21, There was insufficient evidence that avoidance or approach focus longitudinally predicted success, nor that the effect of avoidance or approach focus differed between people (slope variance was not significantly different from zero; p = 0.95). The fixed effects in this model explained less than 3% of the variance in success and the fixed and random effects together explained 10.6% of the variance in success.

Table 21: Multilevel Model Regressing Avoid-Approach on Subjective Success in Study 1

Predictors	*Estimates*	*CI (95%)*
Intercept	2.68 ***	2.34,3.01
Avoid-Approach	0.30	-0.06,0.65
Trait Self-Control	0.27 ***	0.12,0.42
Random Effects		
σ^2	1.79	
τ_{00}	0.24	
τ_{11}	0.11	
ρ_{01}	-0.64	
N	244	
Observations	604	
Marginal R^2 / Conditional R^2	0.029/ 0.106	

Note. CI represents confidence interval. Avoidance-focus was coded as zero and approach-focus was coded as one. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. **p* <0.05 ** *p* <0.01 ****p* <0.001

2.3.6.2 Motivation

Motivation at the beginning of the year was regressed on Subjective Success. As shown in Table 22, Motivation was significantly positively associated with success with within-people and between people, as was Trait Self-Control. There was insufficient evidence that the effect of resolution Motivation varies across people (i.e., that random slopes varied; $p = 0.632$). People felt more subjectively successful in the resolutions that they had more Motivation for initially. Within people, for each one-unit increase in Motivation relative to people's other resolutions, the model predicts a corresponding 0.35 unit increase in success. In addition, people who were more motivated in their resolutions felt more successful at the end of the year. Between people, for each one-unit increase in average resolution Motivation, the model predicts a corresponding 0.26 unit increase in success. The fixed effects in this model explained about 5.1% of variance in success and the fixed and random effects explained about 12.8% of the variance in success.

Table 22: Multilevel Model Regressing Motivation on Subjective Success in Study 1

Predictors	*Estimates*	*CI*
(Intercept)	2.97 ***	2.86,3.09
Motivation (goal)	0.26 *	0.06,0.45
Motivation (person)	0.35 ***	0.15,0.55
Trait Self-Control	0.18 *	0.03,0.33
Random Effects		
σ^2	1.75	
τ_{00}	0.13	
τ_{11}	0.09	
ρ_{01}	-0.54	
N	250	
Observations	636	
Marginal R^2 / Conditional R^2	0.051 / 0.128	

Note. CI represents confidence interval. Motivation (goal) is the group-mean centered Level 1 predictor. Motivation (person) is the grand-mean centered Level 2 predictor. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. *p <0.05 ** p <0.01 ***p <0.001

2.3.6.3 Social commitment

Social Commitment was regressed on Subjective Success. As shown in Table 23, there was insufficient evidence that Social Commitment mid-year longitudinally predicted success between people or within people, accounting for Trait Self-Control. The model with random slopes failed to converge, so a model without random slopes was estimated. The fixed effects in this model explained about 3.4% of variance in success, and the fixed and random effects in this model explained about 11.6% of variance in success.

Table 23: Multilevel Model Regressing Social Commitment on Subjective Success in Study 1

Predictors	*Estimates*	*CI*
Intercept	2.95 ***	2.82,3.07
Social Commitment (goal)	0.05	-0.06,0.16
Social Commitment (person)	0.09	-0.01,0.19
Trait Self-Control	0.26 **	0.10,0.42
Random Effects		
σ^2	1.75	
τ_{00}	0.16	
N	219	
Observations	566	
Marginal R^2 / Conditional R^2	0.034 / 0.116.	

Note. CI represents confidence interval. Social Commitment (goal) is the group-mean centered Level 1 predictor. Social Commitment (person) is the grand-mean centered Level 2 predictor. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. **p* <0.05 ** *p* <0.01 *****p* <0.001

2.3.6.3 Habit formation

Habit Formation was regressed on Subjective Success. As shown in Table 24, fixed effects of Habit Formation both between people and within people were significant, holding Trait Self-Control constant. There was insufficient evidence that slope variances differed from zero (p = 0.17). Within people, for each one-unit increase in Habit Formation, the model predicts a corresponding 0.05 unit increase in success. Between people, for each one-unit increase in Habit Formation, the model predicts a corresponding 0.7 unit increase in success. Fixed factors explained about 8.5% of the variance in success while fixed and random factors explained about 22.9% of variance in success.

Table 24: Multilevel Model Regressing Habit Formation on Subjective Success in Study 1

Predictors	*Estimates*	*CI*
(Intercept)	3.06 ***	2.92,3.20
Habit Formation (goal)	0.05 **	0.02,0.08
Habit Formation (person)	0.07 ***	0.04,0.10
Trait Self-Control	0.19 *	0.01,0.36
Random Effects		
σ^2	1.45	
τ_{00}	0.19	
τ_{11}	0.00	
ρ_{01}	-0.45	
N	175	
Observations	429	
Marginal R^2 / Conditional R^2	0.085 / 0.229	

Note. CI represents confidence interval. Habit Formation (goal) is the group-mean centered Level 1 predictor. Habit Formation (person) is the grand-mean centered Level 2 predictor. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. *$p<0.05$* ** *$p<0.01$* *** *$p<0.001$*.

2.4 Discussion

2.4.1 Research question 1: Characterizing goals

The scope and content of people's New Year's resolutions were varied, but most were consistent with mid-level goals. Many people had more than one resolution. Most resolutions were related to a small subset of life domains: physical health, mental health, and finances, indicated by self-reported nomothetic domain categories as well as by automated word counts of people's resolutions and related open-ended descriptions (i.e., the reasons they gave for setting their resolutions and their descriptions of goal pursuit behaviors). A small subset of goals related to overcoming addiction, often to cigarettes. Human coders rated each resolution on various properties that theory suggests affect success in pursuit. Almost all goals were approach-focused rather than avoid-focused and concrete rather than abstract, and more than half referred to a non-specific or vague endpoint.

People's New Year's resolutions are often more complex than goal pursuits studied in the laboratory (e.g., puzzle completion), but many are similar to goals commonly studied in naturalistic contexts, including other kinds of mid-level goals. For example, many resolutions involved physical health and many naturalistic studies involve exercising. Despite these similarities, New Year's resolutions are unlike other goals typically studied in research in that they are long-term and designed to be achieved in a year rather than a day. This time scale brings with it increase complexity in that pursuit of each resolution is balanced with the pursuit of other goals (and often other resolutions).

2.4.2 Research question 2: Characterizing motivation

People reported consistently high levels of Motivation for their resolutions, as measured by their commitment, effort, and confidence. These three Motivation variables were measured with one item each in January, April, and July. Within time-points (i.e., in January, April, and July), the three variables were positively correlated with one another at levels suggesting that each item captured overlapping but distinct variance in Motivation. It does not seem to be the case that people have low Motivation for their New Year's resolutions, although this study design and the lack of standardization of Motivation measures prevents further assessment of this question.

On average, Motivation declined from January to April and January to July. However, it is not the case that each person got less motivated over time. Although time explained a significant portion of variance in each variable, slopes varied. Resolution trajectories across all three time periods were mostly nonmonotonic. Motivation didn't steadily decline or stay stable for most resolutions. Instead, Motivation seemed to fluctuate. Although measures within the same time period were correlated, even in January when there was little variance, autocorrelations of each measure were low. In addition, the amount of variance accounted for by resolutions and people was modest: 28% to 41%.

Thus, people's levels of effort, commitment, and confidence about their resolutions during the year in part reflect their stable trait Motivation and their stable Motivation for a goal, but there are other sources of variance. Some of this variance may originate in idiosyncratic, dynamic factors that move their Motivation up and down seasonally, and possibly monthly or weekly. These factors could be things like physical health, recent life or cultural events, recent successes in pursuit, or even more transient factors like mood or weather. In our measures of Motivation, some amount of this unexplained variance is not due to actual changes in underlying Motivation, but to our measures and other aspects of our measurement (e.g., how diligent (or not) people were when completing the survey, item phrasing, or self-presentational concerns).

2.4.3 Research question 3: Characterizing pursuit

In both April and July, people described the behaviors associated with pursuit. These followed fairly directly from the content of people's resolutions such that common behaviors related to eating, exercise, and spending behaviors. Many people mentioned behaviors that happen or could happen daily. Prior work

Social Commitment in goal pursuit varied: most commonly, people did not tell others about their goal pursuit. Social Commitment in April and July were moderately correlated (r = .50), suggesting some instability in this construct or its measurement. If the construct changes over time, it should only be in the positive direction (i.e., it's not possible to remove knowledge of the goal from social others). However, Social Commitment means were similar among people's first resolutions in April (M = 3.05, SD = 1.46, N = 311) and July (M = 2.99, SD = 1.46, N = 253). This suggests that the low correlation between waves is due in part to measurement error (e.g., participants forgetting that they'd told people about their resolution).

Habit Formation among people pursuing their New Year's Resolution was moderate. People were above the scale midpoints in both how frequently they pursued their resolutions and how often they pursued their resolutions in the same place, and these two measures – which were only modestly correlated – together constitute Habit Formation. In the sample, Habit Formation was normally distributed, and Habit Formation in April was highly correlated with Habit Formation in July.

It is unclear how typical New Year's resolutions are in terms of Habit Formation and the correlation between frequency of pursuit and context stability of pursuit. Measurement of habits is highly variable (Gardner, 2015). In measures that use frequency and context stability, the correlation between frequency of pursuit and context stability is seldom reported (e.g., Ji & Wood, 2007; Neal et al., 2013). Further, differences between habits in New Year's resolutions and in more narrow goal contexts typical of prior research would likely arise due to differences in goal types alone. New Year's resolutions involve a broad range of goals, but many past studies focus on more narrow behavioral contexts (e.g., transportation habits). For goals that involve behaviors that are highly context-specific, like using a seatbelt, people can vary in the extent to which the behavior is habitual, but overall, the mean value of Habit Formation will be relatively high as will the correlation between frequency and context stability. Other goals, such as eating, are done so frequently and in such varied contexts that Habit Formation will be, on average, relatively moderate, as will the correlation between frequency and context stability. Thus, the most useful comparisons are those with other studies that have measured naturalistic, genuine goal pursuit, and that have used the same or highly similar measures. My own previous work fits this description and my findings in the present study are consistent with my findings in another study that looked at the goals people adopted for Lent (N =

323). In that study, where people's goals were focused on giving up an indulgence (e.g., soda), I found nearly identical values as in this study: the mean of Habit Formation was (M = 14.0, SD = 6.8) and correlation between frequency (M = 3.9, SD = 1.2) and context stability (M = 3.5, SD = 1.3) was r = .40.

2.4.4 Research question 4: Characterizing success

People were about as successful in their pursuit of resolutions in this study as in previous studies of New Year's resolutions. No prior studies have recorded success rates after one year, nor have consistent measures of success been used across studies or in similar timeframes, which makes comparing my results to extant work challenging. Several prior studies have looked at resolution outcomes at three months. In Study 1, 72% of non-missing resolutions were being actively pursued at three months which is higher than the 47% still successful pursuers in a sample of 200 Americans (Norcross & Vangarelli, 1988) and higher than the mean value of 51% success rates at three months in a sample of 254 Americans (Höchli et al., 2019). In Study 1, 39% of people achieved their resolutions at the end of the year, which is slightly lower than the 47%-57% rate based on retrospective reports in large surveys conducted in the UK (e.g., Bupa, 2015). However, the estimated rates of success in this study are compromised to an unknown degree by attrition; the sample at the end of the study was not representative of the entire sample in terms of self-regulatory skill.

In characterizing success, I focused on Subjective Success but measured several outcomes. I found that different measures of outcomes were similar: self-reported achievement was correlated with subjective ratings of success that reflected people's specific barriers and challenges, and people's reported status at the end of the year varied as a function of their success and achievement ratings. These correlations may be attenuated by measurement error, and potentially by systematic measurement error, because aside from Subjective Success, which was explicitly subjective, other measures of goal outcomes may not be invariant across different kinds of goals. For example, the achievement measure probably has a different meaning and differently relates to subjective success for goals that were abstract (and did not have a clear objective criteria for achievement) than for goals that were specific (and did have a clear objective criteria for achievement).

2.4.5 Research question 5: Characterizing disengagement

Deliberate goal disengagement – not pursuing a goal and having no plans to continue to in the future – was rare, amounting to about 5% of resolutions at the end of the year. Further, it tended to be unstable: people who were disengaged from a goal at earlier points later reported being engaged in pursuit. In addition, few people reported explicitly considered giving up on their goals.

Despite the rarity of deliberate disengagement, people seem to be passively disengaging from their goals behaviorally. Throughout the year and at the end of the year – after the ostensible deadline for accomplishing New Year's resolutions – many people reported that their goal was on a temporary hold. In a strict interpretation of New Year's resolutions, any goals not accomplished by the end of the year failed. Yet, these "failed" resolutions remained alive in people's minds: they described most goals that had not succeeded as being actively pursued or on hold rather than deliberately disengaged from. Some unknown portion of resolutions “on hold” will never be resumed again.

For goals that people disengaged from and goals that people put on hold, it might be the case that intentions to resume pursuit, like intentions to start goal pursuits, are not always carried out. The people who reported being disengaged from their goals mid-year and then later reported being engaged in pursuit or achieving their goals demonstrate that people are not rigid in their intentions to *stop* pursuing a goal. Perhaps if I followed up with people who reported putting their goal on hold in another year or two, I would find that intentions to resume pursuit may also not be firm.

Taken together, these findings – and in particular the very low frequency of deliberate disengagement, the unstable nature of disengagement, and the high frequency of passive varieties of goal failure – suggest that theories of goal pursuit and self-regulation mischaracterize goal disengagement as it occurs in daily life in at least three ways. First, persistence and quitting are not the only outcomes of goal pursuit. In this study, the year provided a natural finish line, and at that finish line people often landed between the two – either still plodding along towards their goal (or at least seeing themselves as such) or on a break intended to be temporary. Second, goal disengagement isn’t necessarily permanent. Of the few people who did deliberately disengage from their goals, there was considerable movement to other status categories (e.g., active pursuit, achievement). Finally, contrary to extant theory (Brandstätter & Schüler, 2013; Fishbach & Finkelstein, 2012; Ghassemi et al., 2017; Klinger, 1975), goal disengagement is not always deliberate. Instead, it is possible that people can passively and incidentally disengage from their goals by simply not working towards them (rather than deciding to quit), and indeed, this kind of disengagement is far more common than deliberately giving up.

The findings about goal disengagement in Study 1 highlight the value of descriptive research. Observing these disengagement phenomena in laboratory research and assigned-goal paradigms is unlikely. Even naturalistic study of goal disengagement can be restricted by theoretical assumptions. Our findings are consistent with one prior prospective study of goal disengagement that identified a similarly low rate of deliberate disengagement (Herrmann & Brandstätter, 2015). However, in that study, disengagement was assumed to be deliberate and there was no opportunity to observe passive varieties of disengagement where goals were either “on hold” or seen as active but functionally were inactive (i.e., people weren’t truly pursuing them). In contrast, the passive disengagement I documented in this study is consistent with findings from another descriptive (though not longitudinal) study that documented a similar phenomenon: “frozen goals” that are not being pursued and have not been disengaged from (Davydenko et al., 2019).

2.4.6 Research question 6: Covariation among goal-varying factors

Goal-varying factors in this sample were generally not strongly associated with one another. The one exception to this pattern was that specificity and concrete-abstract variables shared a lot of variance to the point of being redundant. There were also modest relationships between other goal factors. Physical goals were more often concrete and avoidance-focused (although this might be simply due to the prevalence of weight loss goals, substance use goals, and diet goals). The ICCs of goal-varying properties ranged from .00 to .40 and so varied in the extent to which they correlated within people's resolutions. Some factors were more like individual differences and other factors mostly varied at the goal level. Together, covariation analyses suggest that analyses of individual goal-varying factors can be interpreted at face value.

2.4.7 Research question 7: Trait self-control and goal-varying factors

People skilled in self-regulation set more specific goals (as their first resolutions), were more motivated initially in their pursuit, and were more habitual in their goal pursuit. These findings are consistent with prior work that suggests that people with high Trait Self-Control use more effective goal pursuit strategies (Ludwig et al., 2018); that people high in Trait Self-Control often have more and more adaptive varieties of Motivation for their goals (Converse et al., 2019; Milyavskaya et al., 2018); and that people high in Trait Self-Control are more habitual in their goal pursuit (Galla & Duckworth, 2015). Other potential associations between Trait Self-Control that either follow from theory or that have been previously documented did not manifest in this context and with these operationalizations.

2.4.6 Research question 8: Explaining variance in success

Success at the end of the year was associated with self-regulatory skill; people high in Trait Self-Control reported higher levels of Subjective Success in their pursuit at the end of the year. This effect was highly significant, but practically small: the .26-unit increase in success associated with each one-unit increase in Trait Self-Control amounted to an improvement in Subjective Success that was less than 1/5th of a standard deviation. The fixed and random effects in the model (i.e., the fixed effect of Trait Self-Control and the random intercept) explained less than 9% of variance in Subjective Success. Thus, while Trait Self-Control predicts success in New Year's resolutions, to the extent that our measure of Trait Self-Control functions as a proxy for skill in self-regulation, a great deal of variance in success is due to factors other than trait skill.

The majority of goal-varying factors I examined did not predict Subjective Success longitudinally over and above Trait Self-Control. The two factors that did were Motivation at the beginning of the year and Habit Formation in the middle of the year. Both of these effects operated similarly between and within people. Between people, those who started pursuit with more Motivation for their goals (i.e., more commitment, confidence, and effort) felt more successful in their pursuits a year later. Within people, the goals that people had more Motivation for and those that they pursued more habitually were more successful at the end of the year relative to their other goals. Further, there was insufficient evidence that within-person effects varied between people (i.e., that slopes of within-person effects varied between people). The associations of Motivation and Habit Formation with success are consistent with prior literature and theory and offers evidence of the criterion validity of these measures.

Although the effects of Motivation and Habit Formation were highly significant, and the measures I used correspond to different theoretical constructs, the measures I used can also be construed as alternative, and earlier, measures of success in goal pursuit. For example, one of the Motivation items asked about current effort being put towards the goal, and one of the Habit Formation items asked about the frequency of goal pursuit. That early success predicts later success is reassuring but doesn't offer novel insight into the mechanisms of ordinary goal pursuit, nor does it offer useful information about which goals are likely to succeed or fail as resolutions.

My failure to conceptually replicate prior work on goal-varying factors that relate to success might be due to differences in methodology. For example, the measures I used may explain why I didn't find effects. I focused on Subjective Success, but much of the prior research on goal factors has focused on performance outcomes, often in highly measurable tasks (e.g., Weinberg et al., 1985). In addition, the naturalistic context this study was conducted in may have introduced confounds. Further, prior research has tended to control many goal properties that were allowed to vary naturally in this study and may have systematically varied with important variables that I did not measure. For example, Social Commitment (i.e., having told others about a goal) is likely confounded with whether the goal pertains to stigmatized goals or goals that people had strong negative emotions about. It is possible that effects of Social Commitment and other goal-varying properties exist and would emerge if their influence could be cleanly separated from confounds. On the other hand, confounds present in naturalistic work are present in real life. Thus findings that emerge (or fail to emerge) when natural relationships among factors are preserved are more accurate with respect to the real world, and thus more useful, than findings that emerge when factors are made to be independent, whether with statistical adjustments or with study designs. Thus, a reasonable conclusion to reach about the effects of goal-varying properties is that they don't have large enough or universal and unconditional enough influence on Subjective Success to have manifested in this noisy, and yet ecologically valid, goal pursuit setting.

One remaining possibility is that the modeling approach taken here was poorly suited to these data. First, like most other studies of goal pursuit, I may have been incorrect in assuming that goal-varying properties have a linear relationship with Subjective Success. Second, my approach to handling missing data may not have produced unbiased parameter estimates. To understand whether an alternative modeling approach would affect my conclusions, I ran models in a Bayesian framework, as reported in Supplement A. These alternative models support the same basic conclusions as the models reported here. For a few models, the Bayesian analyses could accommodate random slopes while the frequentist analyses could not, and so these models provide some additional information. The models I present in the appendix are just one alternative approach, and so results may differ when using other approaches, such as Bayesian analyses with multiple imputation. However, other approaches, like multiple imputation would offer more compelling evidence that my conclusions are not specific to particular modeling choices.

2.4 Conclusions

The primary aim of Study 1 was to characterize people's pursuit of New Year's resolutions, with a focus on the content and properties of resolutions, characteristics of pursuit, and the fates of resolutions at the end of the year. I found great variety in people's ordinary goal content, properties, and outcomes, which have broader implications for understanding ordinary goal pursuit and goal-varying factors that support success.

The heterogeneity in goal content and characteristics documented in this study suggest that subjective measures of goal success may be the most valid operationalization of successful goal pursuit. People’s goals ranged from overcoming smoking cessation – a notoriously difficult feat often characterized by relapses – to watching every Meryl Streep movie. Success and failure in these two goals are qualitatively different, and likely have different phenomenology and causes. For such a broad variety of goals, more objective outcomes, like frequency of behavioral engagement in pursuit, are ironically *less* valid when used across different kinds of goals that entail different forms and frequencies of pursuit. For example, frequency of behavioral pursuit doesn’t measure success equally well in goals to “work out every day,” to “reduce my level of stress,” or “to lower my blood glucose level.” The kinds of goals people pursue in daily life may not have a clear behavioral manifestation, or may requires different kinds of behaviors, some of which occur routinely and others that are contingent on sporadic events (e.g., temptations, opportunities). For ordinary goals, subjective measures of goal outcomes are likely the most valid kinds of outcome operationalizations.

Although self-reported Subjective Success might be the best operationalization of successful goal pursuit in New Year’s resolutions, other measures of goal outcomes in Study 1 led to novel insights into the nature of ordinary goal pursuit. In Study 1, people reported the status of their goals at each follow up survey. People’s reported statuses throughout the year and particularly at the end of the year make clear that the outcomes of ordinary goal pursuit cannot be simply categorized in terms of success or failure. Some resolutions were achieved, but most were not. Of resolutions that weren’t achieved, a small number had been deliberately quit, but most were in a liminal space between success and failure. Some resolutions were still being actively pursued, some resolutions were put on a temporary hold (or at least a hold intended to be temporary), and others were never started but were intended to be pursued later. The surprising variety and complexity of these outcomes of pursuit have not been observed in the laboratory or short, naturalistic studies.

Several limitations in the method of Study 1 introduce alternative explanations for our many null findings. There was missing data in this study due to attrition, and missingness in the dependent variable limited statistical power to detect effects, and possibly, to accurately estimate the size of effects. In addition, several measures were also implemented in ways that limited their available sample size. Specifically, Goal Domain was measured only in later surveys, and Habit Formation was measured only among people who had reported recently engaging in pursuit.

There were several limitations of Study 1 related to measurement validity and reliability. Human coded goal property variables were based on the text of people's resolutions, which were often just a few words and which often used common or colloquial phrases. This process was resource-intensive, as it required three people to read and assess each resolution. Coders struggled to extract information from people's resolutions and agreement among initial codes was low. In addition to being unreliable, these codes may also not be valid. How people phrased their resolutions may not have represented how people thought about their goals (e.g., in terms of approach or avoidance focus). Another notable measurement issue in Study 1 was related to my measure of

Habit Formation. This measure was based on a popular existing measure (see Gardner, 2015) and borrowed from that measure the unusual step of multiplying two Likert scale responses together. In this sample, the Habit Formation measure had an odd distribution because many of the scale values were impossible (e.g., prime numbers 7, 11, 13, 17, 19, and 23). This odd property of the measure introduces uncertainty about the accuracy and interpretation of regression parameter estimates.

The results of Study 1 suggest that our theoretical models of goal pursuit may oversimplify the effects of goal-varying properties and may oversimplify processes and outcomes of goal pursuit as it occurs in everyday life. Considering limitations related to sample size, attrition, and measurement, this study merits replication and follow-up.

3. Study 2: Predicting Success and Achievement in New Year's Resolutions

The aims of Study 1 were primarily descriptive – to characterize pursuit of New Year's resolution and to glean insights about ordinary goal pursuit. Secondarily, Study 1 offered an opportunity to examine how factors known to affect goal pursuit in more constrained contexts operate in New Year's resolutions. Study 2 replicates and extends the findings of Study 1. Study 2 improves on a few limitations of Study 1 and uses different analytic approaches to better understand the extent to which goal-varying factors can predict success in New Year's resolutions. Specifically, I first report and compare descriptive findings related to the central questions of Study 1. I characterize goals, pursuit, and goal outcomes and assess their relationships with one another, with Trait Self-Control, and with Subjective Success. Then, I turn to the task of assessing the predictive value of regression-based models focused on Subjective Success that use 1) the focal variables from Study 1 as predictors and 2) the focal variables from Study 1 plus all other measured variables that relate to goal-varying properties in this dataset as predictors. Finally, I develop more complex classification models to see how well I can predict Achievement from 1) the focal variables from Study 1 plus all other measured variables that relate to goal-varying properties as predictors and 2) from the subset of the variables in the previous model that were measured at the beginning of the year. This final set of analyses will provide a rough sense for the upper boundary of the predictive value of goal-varying properties as measured in this study.

3.1 Assessing predictive value and predicting goal outcomes

Assessing the predictive value of regression-based models entails more than just estimating regressions and interpreting the amount of variance they explain. Regressions, as they are commonly used in psychological science, are populated with parameter estimates that optimize model fit in the data they are trained in. Typically, in psychology, there is just one dataset per analysis. Models are fit to all observations in that dataset, parameter optimize fit in that dataset, and model performance is assessed in that dataset. When models are trained and tested in the same set of observations, they are likely to be overfit, meaning that they are tailored to those specific observations and leverage idiosyncratic features of those observations (the "error" or "noise" present). When applied to new observations, overfit models perform poorly. For this reason, performance estimates like R^2 that are derived from observations that were used to train a model cannot be trusted to accurately estimate the predictive value of a model (i.e., how well it would work in other observations). As I explain in greater detail below, cross-validation is a machine learning tool developed in computer science that can be used to estimate how well a model would perform in new data and to select among different models on the basis of out-of-sample performance, and in doing so, reduce overfitting.

Outcomes of goal pursuit are multiply determined, and unlikely to be governed by simple linear functions. Developing models that are good at predicting the outcomes of New Year's resolutions likely requires modeling approaches that can accommodate many predictors and that can approximate non-linear relationships. One kind of model that can accommodate many variables, and also reduces overfitting, is penalized regression.

Another kind of model that can accommodate many variables and that can accommodate non-linear relationships between predictors and outcomes are Support Vector Machines (SVM). I provide an overview of cross-validation, penalized regression, and SVM in the following sections. Finally, with these concepts in place, I provide more detail about the benefits of using cross-validation for model selection by comparing it to three other common model selection processes: ad hoc model selection based on sample performance ("*p*-hacking"); systematic model selection based on sample performance; theory-informed, a priori model selection (preregistration); and multiverse analyses that characterize all possible models rather than selecting among them.

3.1.1 Cross-validation overview

Cross-validation is a resampling tool developed in computer science that involves resampling observations and using different sets of observations for model training than for model selection and/or assessment (for a comprehensive introduction, see James et al., 2013). Although cross-validation is relatively new in the psychological sciences, other resampling techniques, like bootstrapping, have been used for decades (e.g., Bollen & Stein, 1993). Cross-validation is kind of resampling that can be used for several purposes, including model selection. For example, in single loop 10-fold cross-validation, a dataset is broken up into 10 unique subsamples of the data which serve as "held-out" validation sets. Then, for each unique held-out subsample, models that differ in their parameter estimates, predictor variables, functional forms, hyperparameters (a class of parameters that adjust machine learning algorithms), and algorithm or model type are trained (i.e., estimated) in the corresponding "held-in" 90% of the dataset. Then, these models – complete with parameter values – are used to predict outcome values in the 10 "held-out" subsamples. K-fold cross validation can split the data into different numbers of subsets, and can be repeated with resampling, such that observations appear in multiple held-out folds. Increasing the number of repetitions of cross-validation is a no-cost way to improve model performance estimates.

This basic process can be used for many purposes, including to select among candidate models. Cross-validation can be used with dozens if not hundreds of different model configurations, and one can be selected based on how well it performs on average in each held-out subsample. Because this selection process optimizes out-of-sample prediction, using cross-validation to select among different models can result in more generalizable and thus more useful and informative models than typical model selection processes in psychology (e.g., selecting a model based on relative performance in an entire dataset; Yarkoni & Westfall, 2017). Cross-validation can also be used to estimate model performance (i.e., by calculating the average performance of a model in held-out folds).

3.1.2 Penalized regressions

Penalized regression models reduce overfitting by intentionally biasing regression parameter estimates. Somewhat counterintuitively, by making parameter estimates slightly *less* tailored to the training set, penalized regressions can improve model fit in new observations. Penalized regressions work by specifically targeting large parameter estimates. Whereas ordinary regressions estimate parameters on the basis of (minimizing) the sum of squared errors, penalized regressions estimate parameters on the basis of a *penalized* function of the sum of squared errors, such that parameters can only be large if they sufficiently improve model fit (for a comprehensive summary of penalized regressions, see Kuhn & Johnson, 2013). There are different kinds of penalized regressions that use different penalties (and thus differently optimize and estimate parameters).

Before I describe three types of penalized regressions, I'll explain an important aspect of many machine learning approaches that has no corollary in modeling approaches typical of psychological science: *hyperparameters.* All penalized regressions use hyperparameters, which, unlike regression parameters, are not estimates of a population effect. Hyperparameters control some aspect of a machine learning algorithm, as I'll later explain in the context of each penalized regression model. Hyperparameter values are best selected via cross-validation (i.e., the values of hyperparameters used in a final model are those that have the lowest cross-validated MSE across held-out folds). Thus, for models that have hyperparameters, the modeling process involves defining a set of candidate hyperparameter values. For algorithms that have multiple hyperparameters, this set will include different combinations of hyperparameter values. Then, models using each candidate hyperparameter value (or each unique combination of hyperparameter values) are trained on held-in folds and tested in held-out folds. This process is often called model tuning. Tuning happens alongside other kinds of model selection such that all candidate models – whether they differ by hyperparameter or by other properties, like the set of predictor variables, the statistical algorithm used, or the functional form of a model – are trained on held-in folds and tested on held-out folds in parallel and the best model among all candidate models (which might potentially vary along many dimensions) is selected on the basis of its average performance across held-out folds.

There are three kinds of penalized regressions commonly used. One kind of penalized regression is ridge regression (Hoerl, 1970). Ridge regression optimizes the sum of squared errors plus the sum of squared regression parameters times lambda, a hyperparameter that controls the amount of penalization. The larger lambda is, the more parameter estimates are penalized, and the more they "shrink" towards zero. Moderately shrinking parameter values with ridge regression often improves out of sample model fit and reduces over fitting.

Another kind of penalized regression is the lasso regression, which imposes a different penalty on parameter estimates. In lasso regression, parameter estimates optimize the sum of squared errors plus the sum of the absolute value of parameter estimates times lambda, which, like in ridge regression, is a hyperparameter that controls penalization. The practical difference between ridge regression and lasso regression is that in lasso regression, parameters can be set to zero and when they are, they are functionally dropped from the model. Consequently, lasso regression selects predictors in addition to reducing overfitting.

A third kind of penalized regression is the elastic net regression (Zou & Hastie, 2005; Friedman et al., 2010) that combines the ridge and lasso penalties. Elastic net regression has two hyperparameters: lambda and alpha. Lambda is the hyperparameter that controls penalization. Alpha controls how ridge-like and how lasso-like the penalization imposed is. When alpha is zero, lambda penalizes as in ridge regression.

When alpha is one, lambda penalizes as in lasso regression. When alpha is between zero and one, the elastic net regression uses both penalties, weighting them differently depending on how close to zero (more ridge) and one (more lasso) alpha is.

3.1.3 Support vector machines

Support Vector Machines are models that were developed in the context of classification (but can also be applied in a regression framework; for a thorough introduction see James et al., 2013). In classifying outcomes, SVM seeks to define a boundary between classes in multidimensional space. In a simple case of two classes that can be well characterized by two predictors, for example, each point can be positioned in two-dimensional space (i.e., the familiar cartesian plane) and a straight line can be placed between the classes such that all observations below the line are in one class and all observations above the line are in another. New observations are predicted to belong to a class based on where they fall relative to the line. When more variables are included and more dimensions are used to characterize the two groups, the line becomes a hyperplane.

There are different ways that an algorithm can position a classifying hyperplane. In a simple case, a classifier can define the hyperplane that maximizes the margin (i.e., distance) between the hyperplane and all observations. In this approach, points near the hyperplane have a large influence on it and can result in overfitting. In addition, many groups are not cleanly separable by a hyperplane. Thus, more useful approaches to defining a boundary between classes allow some points to be close to the hyperplane and even on the wrong side of the hyperplane. Even in cases where classes can be perfectly separated, just as penalized regression can produce better out-of-sample estimates than regression that maximizes prediction in the training set, a hyperplane that doesn't perfectly classify all observations in the training set can produce better out-of-sample prediction.

This basic aim and logic of optimally defining a hyperplane between two classes is generalized to non-linear decision boundaries in SVM. The math of the SVM algorithm is complex. It relies on an estimation of the similarity of two observations, conceptually similar to Euclidean distance, using a function called a *polynomial kernel*. Kernels can be used to optimize and define non-linear, and even circular class boundaries (using a *radial kernel*). SVMs can accommodate a large number of predictors and complex relationships among predictors and between predictors and outcomes. Hyperparameters in SVM are a cost hyperparameter C, that regulates the model's tolerance for observations near or on the wrong side of the hyperplane, and the Sigma hyperparameter that controls how sensitive the model is to observations far away from the hyperplane. SVM can classify observations into two groups better than many other classification approaches (e.g., Linear Discriminant Analysis). However, it is computationally infeasible (and in many cases impossible) to characterize the influence of predictor variables. Thus, predictive capacity of SVM comes at the cost of useful inference about how the model is making predictions.

3.1.4 Comparing model selection in cross-validation to other model selection approaches

Cross-validation can be used to select among models that differ along many different dimensions. Often, cross-validation is used in machine learning to select among models that use different statistical algorithms, that have hyperparameter values, or that use different predictors or sets of predictors.

There are four common model selection procedures used in psychology, which range in their formality and in what they optimize. First, models are sometimes selected for a set of ad hoc candidate models based on their performance in a sample ("*p*-hacking"). In this approach, various versions of a model are run, for example, with different covariates or interaction terms, often unsystematically. Then, a model is selected from among these varieties based on performance in the entire sample (or a function of performance, like obtaining a *p*-value below 0.05). This model selection process optimizes fit in the sample, which can lead to overfitting and can produce models that may not provide accurate estimates of population parameters and that may not perform well in new observations, depending on how many different models were tested (Simmons et al., 2011).

More often, models are selected on a more principled basis. For example, formal model comparison procedures are often used to compare small sets or pairs of models on the basis of their performance in a sample. For example, hierarchical regressions are often used to compare fit among models that differ with respect to the inclusion of predictors (e.g., interaction terms), and SEM is used to compare fit among models that differ in parameter constraints or other aspects of model specification. This model selection process optimizes fit in the sample and consequently, can lead to overfitting, too, especially in small samples (Yarkoni & Westfall, 2017). Another kind of model comparison – not often framed as such – involves selecting a model a priori based on theory and specifying this model in a preregistration. This model selection process doesn't optimize anything, mathematically, and thus doesn't result in overfitting. A less common variety of model selection doesn't select one model per se, but instead explores the analytic space using, for example, a multiverse analysis (Steegen et al., 2016) or specification curve analysis (Simonsohn, Simmons, & Nelson, 2015). In this approach, all or many possible variations of a model are estimated, and the modeling space is characterized such that the effects of different dimensions of analytic choices on results can be understood (e.g., Rohrer, Egloff, & Schmukle, 2017; Orben & Przybylski, 2019).

In sum, cross-validation is a useful tool for model selection that combines elements of extant model selection procedures. Like ad hoc model selection ("*p*-hacking"), cross-validation allows for data exploration and optimization of model performance. Crucially, though, cross-validation can be used to optimize out of sample performance and thus can reduce overfitting. Like formal model comparison procedures in the context of regression or SEM, cross-validation can be used to compare theoretically meaningful sets of predictors, and select among them on the basis of model fit, but in a way that reduces overfitting and is more likely to generalize beyond a particular set of observations. Like a priori model selection with preregistration, cross-validation can reduce analytic flexibility and be used to produce a model that has not been tailored to a set of observations and less likely to lead to false positives. Finally, like multiverse analyses, cross-validation can provide information about how analytic choices affect model fit. Cross-validation generates information about the fit of every candidate model, although unlike in many multiverse analyses, fit is evaluated in held-out folds rather than in the entire sample. Cross-validation is a useful tool that combines many beneficial aspects of extant model selection procedures, without some of their drawbacks, allowing for robust model selection with flexible exploration. In this Study, I use single k-fold cross-validation to estimate the performance of ordinary regression, and to train, tune, and select a penalized regression model and two SVM models. This use of cross-validation is mostly for demonstration purposes rather than for obtaining the most precise estimate of out-of-sample performance possible. If my goal was to obtain the most accurate estimate of out-of-sample performance, I would use an alternative approach to estimating the performance of the final model selected with k-fold cross validation, like a hold-out sample or nested cross-validation (Ding et al., 2014; Guan, Xiang, Deng Keating, 2004; Jonathan, Krzanowski, McCarthy, 1999)

3.2 Method

Study 2 was broadly similar to Study 1 but with a simpler design that was squarely focused on the study aims. Sufficiently complex designs reduce analytic flexibility and participant burden. Whereas Study 1 was designed as an initial exploratory study focused on the reasons that people disengaged from their goals, Study 2 was designed as a follow-up to Study 1 that could replicate and extend the findings in Study 1. There were twenty or so variables measured in Study 1 that were not particularly useful to analysis, mostly because they had near zero variance due to my incorrect assumption that the majority of participants would give up their resolution during the year. A few variables were simply included on a whim and were atheoretical and uninformative, often because they were poorly measured (e.g., whether people were using any mobile phone apps to help them pursue their goal). I measured far fewer variables in Study 2. In addition, the multiple mid-year time points captured in Study 1 were not, and could not easily be leveraged in analyses, and so the design was simplified to one mid-year time-point. Thus, the primary difference between Study 1 and Study 2 was that there were just three time-points (January, July, and January) and fewer quantitative measures. Several measures were adjusted: rather than using human coders, participants self-reported goal qualities and the poles of goal qualities were rated separately rather than assumed to be orthogonal (e.g., goals could be rated as involving approaching a desired state and avoiding an undesired state). Finally, participants reported Subjective Success as in Study 1, but they reported Achievement as binary variable. Binary outcomes are conceptually useful, and useful for predictive algorithms. Study 1 made clear that a binary achievement outcome could not easily be derived from, for example, participants' reported status, given how many people ended up somewhere between success and failure at the end of the year. Thus, in Study 2, I asked participants to provide a binary assessment of their own achievement. In addition, Study 2 has a larger sample size and more statistical power than Study 1.

3.2.1 Sample

Participants were native English speakers in the United States, recruited with Amazon Mturk (M_{age} = 37.1, SD_{age} = 12.0). There were 339 male-identified people in the sample, 467 female-identified people, and 3 people who were nonbinary, preferred not to say, or preferred to self-describe. People who participated in Study 1 were unable to participate in Study 2. Participants were excluded from analyses if their open-ended responses were insincere or suspicious. In addition, 23 people admitted to answering questions randomly in one or more surveys and their data were excluded from analyses.

Participants were well-educated, with rates almost identical to Study 1. Over 99% of the sample had a high school degree or more. The mode of educational attainment was a Bachelor’s degree (37%; n = 288). Many people had completed some college (26%; n = 205) or an Associate’s degree (15%, n = 118). About 14% of the sample had completed or worked towards a graduate or professional degree (n = 111).

The sample was mostly White and indicated their ethnicity as mostly not Hispanic or Latino (n = 736; 91%). Participants’ non-exclusive racial identities were as follows: n = 670, 82% White; n = 74, 9% Black; n = 55, 7% Asian; n = 13, 1% Native American or Alaska Native; n = 2, <1% Hawaiian or Pacific Islander; or another identity, n = 17, 2%.

2.2.2 Procedure

In mid-January of 2018, participants were invited to participate in a survey, which was described as requiring that people had set New Year’s resolutions. The first question participants saw was a screener that asked how many resolutions participants had set. If they set more than one, they were invited to participate and given informed consent. Those who agreed to participate then began the survey which entailed completing individual difference measures, reporting up to five resolutions – which they were instructed to list in order of importance – and answering questions about each one. Participants were invited to participate in additional surveys in mid-July (T2), and January of 2019 (T3) via TurkPrime. Participants had approximately two weeks to complete each survey and were sent up to two additional invitations.

Payment for the first survey was $1.50 and it took participants an average of 19.7 minutes to complete. Payment for the second survey was $2.00 and it took participants an average of 16.3 minutes to complete. Payment for the third survey was $3.00 and it took participants an average of 9.7 minutes to complete.

Surveys as administered and data are available at osf.io/muar5.

3.2.2 Measures

Measures used in inferential analyses are described here. Measures used in descriptive analyses are reported in the results, and descriptive statistics for the predictors included in the exploratory and predictive analyses are reported in Supplement B.

In this study, participants rated their own goals in terms of concreteness, abstractness, approach-focus, and avoidance-focus, with measures created for this study. Concreteness was measured as agreement on a 5-point Likert scale with the statement "This resolution is concrete and objective. I will know for sure when I have succeeded" from 1 (*not at all*) to 5 (*completely*). Abstractness was measured as agreement on the same scale with the statement "This resolution is abstract and subjective. I will have a general sense of my success." Approach-focus was measured as agreement on the same 5-point Likert scale with the statement "This resolution involves starting or approaching something." Avoidance-focus was measured as agreement with the statement "This resolution involves stopping or avoiding something."

Motivation was measured in January (T1) and April (T2) with the same three items as in Study 1, measuring commitment, effort, and confidence on 5-point Likert-scales. Reliability of the three items together (assessed with one resolution per person) was high at T1 (α =.71, ω = .73) and T2 (α =.86, ω = .87). At T1, responses were skewed which likely attenuated the composite reliability.

Social Commitment was measured in January (T1) with one item written for this study that asked participants to indicate whether they had told others about their resolution ("I have told other people about this resolution"). Participants indicated how true of them the statement was on a Likert-scale from 1 (*not at all true*) to 5 (*completely true*).

Habit Formation was measured in July (T2) with the 4-item Self-Report Behavioral Automaticity Index (SRBAI; Gardner et al., 2012). The scale contains items that asked about perceived automaticity of behaviors (e.g., "[Behavior] is something I do without having to consciously remember"). Participants indicated their responses to each item on a Likert-scale from 1 (*strongly disagree)* to 8 (*strongly agree*). Reliability was high in July (e.g., within people's first resolution, α =.95, ω = .96)

Subjective Success in goal pursuit was measured at the end of the year (T3) with one item written for this study. Participants reported the extent to which they felt successful in a subjective sense, considering modifications to their goal and considering constraints on a Likert-scale from 1 (*not at all*) to 5 (*completely*).

Achievement was measured in July (T2) and at the end of the year (T3) with one item written for this study. Participants reported whether "In a literal sense, have you achieved this resolution?" selecting either "yes" or "no."

As in Study 1, Trait Self-Control was measured with the 20-item Capacity for Self-Control Scale (Hoyle & Davisson, 2018). The scale had excellent reliability (α = .93;

ω = .95). In this sample, Trait Self-Control was associated with BFI-44 Conscientiousness (α =.89, r = .76).

3.2.3 Analytic approach

In revisiting the questions of Study 1, I assess the content of goals, characterize pursuit, and the prevalence of success and goal disengagement in all resolutions. I present descriptive statistics and correlations among people's first resolutions for the focal goal-varying properties: Goal Domain, Concreteness, Abstractness, Approach, Avoidance, Motivation, Social Commitment, and Habit Formation. I also present correlations of these factors with Trait Self-Control and Subjective Success. In Appendix B, I present 1) descriptive statistics and correlations derived from the entire sample and 2) frequentist and Bayesian multi-level models regressing each factor on Subjective Success.

To assess the predictive value of regression models using the focal goal-varying factors and Trait Self-Control to predict Subjective Success, I use linear (gaussian) regression in the R package caret with cross-validation (Kuhn, 2008). I restrict these analyses to people's first resolutions, which simplifies the modeling process. Just as in the context of regression models trained and assessed on a single sample, restricting analyses to just people's first resolutions reduce statistical power and represents a loss of information. However, this analytic choice is a pragmatic solution and is consistent with the aim of assessing the effects of goal-varying factors that theory suggests are universal.

To assess how well a larger set of variables explain variance in Subjective Success, I use gaussian elastic net regression with cross-validation in caret. Values of hyperparameters will be selected via cross-validation from a 3x3 grid crossing three default values of alpha and three default values of lambda. I include in this model many variables that were measured in this study but that either were exploratory or had no corresponding measures in Study 1. These variables and basic descriptive information about them are reported in the Results section.

To develop good predictive models that can predict Achievement, I use a radial kernel support vector machine (SVM) in caret using data from all resolutions. The values of the SVM hyperparameters will be selected via cross-validation from a tuning grid of ten combinations of a default-selected value of Sigma and 10 values of C. In order to better understand what information the SVM made use of, and to understand how well a model might be able to predict achievement using temporally distal goal-varying variables, I will repeat this process using only variables collected at the beginning of the year. The same procedure will be used for the model that include all predictors and the model that includes only those predictors measured at the beginning of the year.

For each analysis using cross-validation, I use five repeats of 10-fold cross-validation. I follow standard practices for preprocessing predictors in machine learning (e.g., as described in Kuhn & Johnson, 2013): I remove variables with near-zero variance, identify highly multicollinear predictors (and remove redundant variables), center and standardize all predictors, and used median imputation for predictors with missing values. I use listwise deletion for missingness on the outcome variable.

3.3 Results

3.3.1 Characterizing goals, pursuit, success, and disengagement

The modal number of resolutions (47%) was one ($M_{resolutions} = 1.72$). In this study, participants were first asked the number of resolutions they had and later were asked to report what their resolutions were. One hundred and fifteen people listed more resolutions than they had initially reported and were asked follow-up questions about only as many resolutions as they had first reported.

As shown in Table 25, the most common word stems in people's resolutions related to physical health. An estimated 7% (n = 99) of resolutions pertained to substance use. The most common bigrams in people's resolutions were "lose weight" (n = 136), "eat healthier" (n = 39), "quit smoking" (n = 31), and "save money" (n = 28).

Table 25: Top 10 Word Stems in New Year's Resolutions and Reasons in Study 2

	Resolutions		**Reasons**	
Rank	Word	*n*	Word	*n*
1	Lose	241	Feel	552
2	Weight	202	Set	439
3	Eat	109	Resolution	435
4	Exercise	81	Weight	399
5	Money	76	Time	396
6	Pound	76	Life	316
7	Save	74	Health	307
8	Time	69	Money	244
9	Healthier	61	Eat	239
10	Smoke	61	Goal	236

Note. Stop words were removed and all words were stemmed. Word stems were edited for readability (e.g., "exercis" as "exercise").

All participants reported the non-exclusive life domains their resolution related to. The most common domains were Physical (*n* = 822) and Mental (*n* = 645). As shown in Table 26, considering combinations of selected domains, the most common pattern was physical health as the only selected domain, followed by physical health and mental health, and then money as the only selected domain. Of the 27 resolutions that were associated with a domain not listed, the words stems that appeared more than once across people's open-ended domain descriptions were: "hobby" (*n* = 4), "health" (*n* = 3), "love" (*n* = 2), "relationship" (*n* = 2), and "time" (*n* = 2).

Table 26: Frequencies of Patterns in Life Domains Associated with at Least 20 New Year's Resolutions in Study 2

Rank	Physical	Mental	Money	Career	Social	*n*
1	✓					350
2	✓	✓				187
3			✓			125
4		✓				54
5			✓	✓		33

Note. Among patterns of life domains associated with at least 25 resolutions, none related to the five other domain options: Family, Society, Spiritual, Education and Other.

Goal-varying properties are described in Table 27. Among people's first resolutions, more resolutions pertained to Physical Domain (*n* = 551) than did not (*n* = 243). Fewer resolutions pertained to Mental Domain (*n* = 362) than did not (*n* = 432).

Table 27: Means, Standard Deviations, N, and Correlations of Goal-Varying Properties Derived from People's First New Year's Resolution in Study 2

Variable	*M*	*SD*	*N*	1	2	3	4	5	6	7	8
1. Physical Domain	0.31	0.46	779								
2. Mental Domain	0.45	0.50	779	-.08*							
3. Concrete	4.26	1.04	777	-.08*	-.08*						
4. Abstract	2.49	1.50	775	.04	.09**	-.59**					
5. Approach	4.06	1.23	778	.11**	.04	.04	.03				
6. Avoid	3.52	1.58	776	-.33**	.01	.07	.07	-.30**			
7. Motivation	4.42	0.61	779	.04	.03	.20**	-.01	.11**	.01		
8. Habit Formation	3.57	1.99	440	.05	.06	-.07	.11*	.17**	-.08	.20**	
9. Social Commitment	3.48	1.51	778	-.05	-.02	.15**	.00	.03	.08*	.25**	.05

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. *N* is used to indicate the available sample size for each variable. Habit Formation was derived from the survey administered in July (T2). All other variables were derived from the survey administered in January (T1). Binary variables list the level coded as zero first and the level coded as one second. * indicates $p < .05$. ** indicates $p < .01$.

As shown in Table 27, most resolutions were high in concreteness and below the mid-point on average in abstractness. Concreteness and abstractness were negatively correlated at around $r = -.6$, suggesting that they are related but somewhat separable dimensions of goal pursuit. On average, resolutions were highly approach-focused, but also above the mid-point on avoidance-focus. Approach and avoidance were negatively associated at $r = -.3$ suggesting that approach and avoidance are non-exclusive and separable. Motivation at the beginning of the year was very high. As shown in Figure 12, motivation lowered on average from January to July, but there was variation in resolution trajectories. Habit Formation was moderate (out of a maximum score of 7, resolutions were on average just below the mid-point of four). Social Commitment was above the mid-point, suggesting that the majority of resolutions were known to people other than the pursuer.

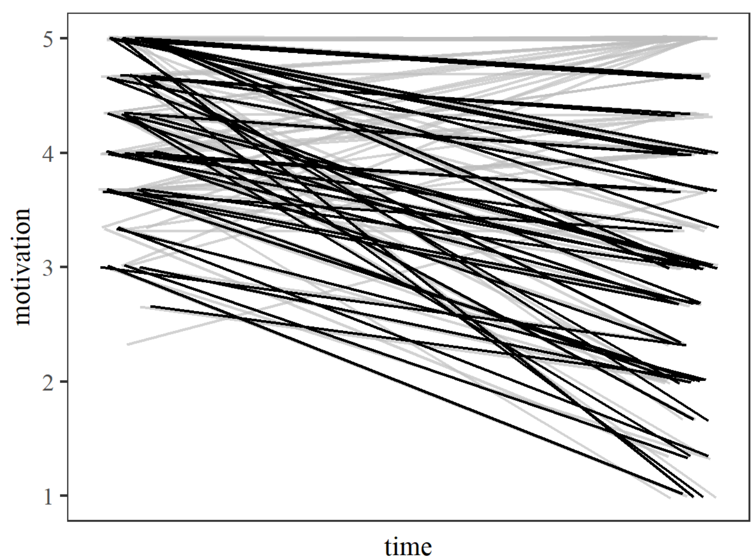

Figure 12: Motivation Trajectories of 100 Randomly Selected Resolutions from January (T1) to July (T2) in Study 2 with Decreasing Trajectories Highlighted

As shown in Table 28, ICCs ranged from 0 to 0.42. Goal-varying properties were varied in the extent to which they varied between people, with Social Commitment, Motivation, and Abstractness at the high end (i.e., people's resolutions tended to have similar levels of Motivation and Abstractness) and with Habit Formation and Avoidance at the low end (i.e., people's resolutions were not highly similar with respect to Habit Formation and Avoidance).

Table 28: ICCs and Number of People and Observations for Goal-Varying Properties in Study 2

Variable	*ICC*	*N* people	*N* resolutions
Physical Domain	0*	794	1360
Mental Domain	0.23	794	1360
Concrete	0.21	792	1353
Abstract	0.42	789	1352
Approach	0.20	793	1358
Avoid	0.17	791	1356
Motivation	0.42	794	1360
Habit Formation	0.16	447	774
Social Commitment	0.50	793	1358

Note. *ICC* represent the intraclass correlation estimated from a model with a random intercept for people. * indicates that the estimated ICC was zero because the estimated random intercept variance was zero.

In the middle of the year, of people who completed the survey, relatively few were successful. Just 2% (n = 16) of resolutions were achieved and done, and 17% (n = 133) were achieved and being maintained. Most resolutions (56%) were described as actively being pursued (n = 433). 16% were described as on hold (n = 126). 4% (n = 31) were disengaged from and 4% were not yet started (n = 27). Among the 184 people who had stopped pursuing their goal in the middle of the year (people who had put their goal on hold, had disengaged, or had not started) most people reported that their decision was not deliberate (61%, $n = 113$). In open-ended text descriptions of the reasons that people had stopped pursuit, the most common word was "time" (appearing in 39% of resolutions; $n = 72$).

At the end of the year of people who completed the survey, about 17% of resolutions were achieved, either achieved and done ($n = 25$) or achieved and maintaining ($n = 105$). Most resolutions (62%) were described as actively being pursued ($n = 463$). The rest were not being pursued and were either never started (3%; $n = 23$), disengaged from (3%; $n = 24$), or on hold (15%; $n = 113$).

People's self-reported Subjective Success and Achievement offer a slightly different (though broadly similar) depiction of goal outcomes at the end of the year. People felt moderately successful; the modal response was the scale midpoint ($M = 2.92$, $SD = 1.29$). People reported an achievement rate of 40% ($n = 305$).

As shown in Table 29, the majority of goal-varying properties were associated with Trait Self-Control. People higher in Trait Self-Control tended to have (first resolutions) that were more Concrete, less Abstract, that they had more Motivation for, that they pursued more automatically, and that they told others about more than those who scored lower in Trait Self-Control.

Table 29: Correlations of Each Goal-Varying Property with Trait Self-Control Derived from People's First Resolutions in Study 2

Variable	*r*	CI
Physical Domain	-0.07	-.14, .00
Mental Domain	-0.01	-.08, .06
Concrete	.13**	.06, .19
Abstract	-0.06	-.13, .01
Approach-Focus	0.05	-.02, .12
Avoidance-Focus	-0.03	-.10, .04
Motivation	.31**	.25, .37
Habit Formation	.11*	.02, .20
Social Commitment	.13**	.06, .20

Note. Trait Self-Control $M = 3.55$, $SD = 0.73$, $N = 794$.

As shown in Table 30, only Motivation and Habit Formation were associated with Subjective Success. Among people's first resolutions, resolutions that had higher initial motivation and those that people pursued more automatically were those that people felt more successful in at the end of the year.

Table 30: Correlations of Each Goal-Varying Property with Success Derived from People's First Resolutions in Study 2

Variable	*r*	CI
Physical Domain	0.05	-.05, .14
Mental Domain	0.01	-.08, .10
Concrete	0.03	-.07, .12
Abstract	-0.07	-.16, .02
Approach-Focus	0.03	-.06, .12
Avoidance-Focus	-0.08	-.17, .01
Motivation	.20**	.11, .28
Habit Formation	.33**	.23, .42
Social Commitment	0.05	-.04, .14

Note. Subjective Success *M* = 3.01, *SD* = 1.27, *N* = 448.

3.3.2 Assessing regression analyses predicting success

A multiple linear regression using the nine focal goal-varying factors and Trait Self-Control had a cross-validated R^2 of 0.105. The final model parameter coefficients are listed in Table 31. Variable importance – which is a function of the t-statistic of regression parameters – are shown in Table 32. The regression mostly relied on information about Habit Formation, initial Motivation, Abstractness, Avoidance-Focus, Approach-Focus, and Concreteness. However, the model explained just over 10% of the variance in held out folds.

Table 31: Parameter Coefficients from Cross-Validated Linear Regression in Study 2

Parameter	Coefficient
(Intercept)	3.009
Physical	0.029
Mental	0.014
Abstract	-0.144
Concrete	-0.068
Approach	-0.077
Avoid	-0.084
Motivation January	0.200
Habit Formation	0.390
Social Commitment	0.035
Trait Self-Control	-0.001

Table 32: Predictor Importance from Cross-Validated Linear Regression in Study 2

Predictor	Importance
Habit Formation	100.00
Motivation January	55.15
Abstract	33.96
Avoid	20.52
Approach	18.54
Concrete	16.54
Social Commitment	8.57
Physical	6.30
Mental	0.45
Trait Self-Control	0.00

3.3.3 Exploratory regression analyses predicting success

An elastic net regression using a relatively large set of predictors, in addition to the focal goal-varying factors, had a cross-validated R^2 of 0.24. The hyperparameter values that defined this model were Lambda = 0.214 and Alpha = .6. The model imposed more ridge penalty than lasso penalty. The predictors added to the model included all life domains (except other and society, which had near-zero variance), people's use of pursuit strategies (e.g., planning how, planning when, tracking progress), an assessment of goal value, whether resolutions were new, various costs of pursuit, whether people were able to recall their resolution in July (Forgot July), July Motivation, and whether people had achieved their resolution in July (Achieved July). The top 10 variables, ranked by importance – a function of coefficient size – to the final model are listed in Table 33. The regression mostly relied on Progress in July, Motivation in July, and Achievement in July. The model explained a middling 24% of the variance in held out folds.

Table 33: Predictor Importance of Top 10 Predictors from Cross-Validated Elastic Net Regression in Study 2

Variable	Importance
Progress July	100.00
Motivation July	62.09
Achievement July	35.57
Track Progress	4.71
Modified July	1.48
Progress January	1.41
Abstract	1.13
Status-On Hold	0
Status-Achieved	0
Physical	0

3.3.4 Predicting achievement

An SVM classifier using the same set of predictors as the elastic net regression had a cross-validated AUC of 0.735 and an accuracy across all predicted and observed values of .67 [.6685,.6779], a sensitivity of .84, and a specificity of 0.42. It was characterized by the hyperparameters C = 0.25 and Sigma = 0.017. This model was an improvement on a null model that used only class probabilities to make predictions at an accuracy of 0.60 ($p < .001$). The SVM with all variables, including a prior measure of the outcome six months earlier, only slightly improved on null model (at best a 6.8% improvement in classification accuracy).

An SVM classifier using only predictors collected at the beginning of the year had a cross-validated AUC of 0.67 and an accuracy in the entire dataset of .63 [.6256,.6353], with a sensitivity of .87 and a specificity of 0.27. The model was characterized by the hyperparameters C = 0.25 and Sigma = 0.026. This model was an improvement on a null model that used only class probabilities to make predictions at an accuracy of 0.60 ($p < .001$). The SVM with variables measured only at the beginning of the year was worse than the model with information from July, suggesting that the previous model – like the regressions modeled before it – were relying on information from July. This SVM offered very little improvement over the null model (at best a 3.5% improvement in classification accuracy).

3.4 Discussion

3.4.1 Updating the conclusions of Study 1

Goal content was nearly identical across studies in terms of words in resolutions and life domains, although more resolutions were related to mental health in January of 2018 than in January of 2016. Of people who selected a domain not listed, the only common domain across Study 1 and Study 2 was “hobby.” The five most common patterns of domains were nearly identical in rank and relative frequencies across studies. In Study 2, I again found that most resolutions were approach-focused and concrete, despite using a different measure (coder-rated categorically in Study 1, self-rated continuously in Study 2).

People had fewer resolutions in Study 2 than in Study 1, which owes to differences in how resolutions were solicited. In Study 2, people were first asked how many resolutions they had, and later were asked to report the content of their resolutions (up to five, ordered in terms of importance). People reported fewer resolutions than they listed. When it came time to describe their resolutions, they often had more than they thought (or, more likely, they were often inspired to set new resolutions). In Study 1, participants were surveyed on all the resolutions that they described, but in Study 2, participants were surveyed only on the number of resolutions and were only asked follow-up questions about the number they explicitly reported. Consequently, people had more resolutions in Study 1 than Study 2, and in both studies, people seem to have used the survey to set goals, not just report on goals they were already pursuing.

Other than this difference in the number of goals, success outcomes, disengagement, and the descriptive analyses and covariations among goal-varying factors were broadly similar across each study, despite some small differences that likely relate to measurement. In both studies, I found that deliberate disengagement was rare and that many resolutions were still active or were on a temporary break at the end of the year. In Study 2, relatively more people reported their resolutions as being active rather than achieved. In addition, achievement as measured by reported status was lower in Study 2

(20%) than Study 1 (40%). In Study 2, participants were asked to first reflect on their status verbally and were also asked to provide a rating of binary achievement, which may have affected how they answered the status question. Curiously, the binary achievement rate matched the 40% achievement status rate reported in Study 1. About 15% of people reported that their goal was on hold at the end of the year. Earlier analysis of status shows that of people who had stopped pursuing their goal in July (about 24% of non-missing observations), most had not deliberately decided to. The single most common word among open-ended explanations for stopping was "time."

Most goal-varying factors were unrelated to others, except that Social Commitment, Habit Formation, and Motivation, covaried, and the latter two related to Trait Self-Control and Subjective Success.

3.4.2 Assessing regression analyses

A regression analysis using the focal goal-varying properties had a modest out-of-sample performance of R^2 of 0.105. Consistent with the results of Study 1 and the supplemental analyses for Study 2 (in Appendix B), the model mostly made use of Habit Formation and Motivation. Trait Self-Control was not used by the model, but recall that it covaried with Habit Formation and Motivation, and that linear regression parameter estimate the unique influence of each predictor over and above covariation with other predictors. The cross-validated estimate of total variance explained is less biased than an estimate produced by simply training a regression on the entire sample and measuring its performance, but it may still be a biased estimate of the true model performance in new data. Performance estimates from singe k-fold cross-validation are averages and can be influenced by a fluke high performance in one or more held-out folds. Cross-validation can only select the *best* model among a set of models; if a good model isn't present in the set of candidates, cross-validation cannot produce one. Most likely, a linear regression with a relatively small set of variables is not complex enough for predicting Subjective Success in ordinary goal pursuit, and so none of the candidate models that I considered within cross-validation performed well.

3.4.3 Exploratory regression analyses

A penalized regression analysis using the focal goal-varying properties plus dozens of other predictors fared slightly better in predicting Subjective Success than the previous model and had an out-of-sample performance of $R^2 = 0.24$. Although linear regression is likely still too simple to meaningfully explain variance in Subjective Success, including more predictors and reducing overfitting with penalization improved model performance. The variables that were most important to the elastic net regression were proximal motivation and goal performance measures from July, including an earlier measure of the outcome, Achievement. Other important variables related to initial progress and strategy use. The only goal-varying property that was important to the exploratory regression was goal abstractness, which likely explaining variance in measurement error in Achievement rather than true differences in Achievement as a function of abstractness.

3.4.4 Predicting achievement

Predictive models are not well-suited for explaining mechanisms or for theoretical inference. They do provide information about the predictive signal of variables. An SVM using a set of predictor variables that is relatively large in goals and self-regulation research had mediocre performance of an estimated 67% accuracy (i.e., 67% of cases were correctly classified). When restricting a predictive model to information available at the time of goal setting – a year before the key outcome – the model performed much worse at 63% accuracy, although still slightly better than a model using just the prevalence of goal achievement. Although cross-validation can provide much better estimates of model performance than standard approaches in psychology, estimates derived from k-fold cross validation are often positively biased and the model would likely perform slightly worse in truly new data or a true hold-out set.

It is probably the case that there is a ceiling on how well any model using goal-varying factors, and even measures of goal progress mid-year, could predict goal pursuit in daily life because of the inherent vulnerability of goals to stochastic events. Goal pursuit in daily life can often be derailed by unrelated factors that no person would be able to anticipate, let alone be able to provide relevant information about via self-report. These events could be things like negative interpersonal conflicts, the behavior or needs of others, including children, physical illness, or financial losses. Some people are probably more vulnerable to uncontrollable and unpredictable events that might derail their goal pursuit (e.g., people living in poverty). There is no theoretical or common-sense reason that differences in vulnerability to unexpected goal derailment would be meaningfully related to goal-varying factors like goal specificity that people might be able to strategically control.

With the variables used in these models, the prediction ceiling seems to be about 67% accuracy. There might be ways to improve on the accuracy level in this dataset, for example by taking full advantage of the open-ended text that people provided or by theory-informed predictor engineering (e.g., including interactions). In addition, although radial kernel SVM often performs very well relative to other learning algorithms (and it performed better than elastic net regression in this dataset), using a broader range of predictive algorithms would likely improve on the best accuracy rate. Providing cross-validation with more models increases the chance of identifying a good or even great model.

In addition, there are probably ways to improve on this rate in new datasets, by collecting additional information, including theoretical goal-pursuit variables not examined here and variables that might index people's vulnerability to external, disruptive events. Further, although cross-validation and the algorithms used here can accommodate complex relationships and large numbers of variables, they cannot overcome measurement error in predictors or especially in outcomes. Obtaining better measures – or even restricting a study to goals of one type in order to reduce systematic measurement error in Achievement and Subjective Success confounded with goal type – would likely raise the ceiling, too. However, the more a set of goals is restricted, the less general the model and any inferences that could be gleaned from it, would be.

4. Conclusion

This dissertation aims to advance our understanding of theoretical, goal-varying factors that meaningfully affect goal pursuit in daily life. It first aimed to characterize ordinary goal pursuit and then aimed to identify goal-varying factors that predicted Subjective Success and to estimate how well goal-varying factors could predict Achievement longitudinally. In accomplishing these aims, this dissertation adopted an observational method where people reported on several goals in ways that I hoped could disentangle within and between person effects. In addition, I used analytic tools form machine learning to estimate and optimize prediction, with the hope that doing so would elucidate how well goal-varying factors can predict goal outcomes and might identify specific factors that do meaningfully predict goal outcomes. In this concluding chapter, I assess each aim in turn, noting the benefits and challenges of the methodological and analytic approaches I took. I then reflect on the current state of research related to effective goal pursuit. I end with a call for more naturalistic research and more research focused on describing and predicting real-world outcomes rather than on developing and elaborating on explanatory theory.

4.1 What is ordinary goal pursuit like, and what are typical fates of ordinary goals?

Ordinary goal pursuit as characterized in this dissertation resists simple characterization: the goals people set related to the same basic domains but were extremely varied in their content and how they were phrased. Most goals were approach-focused and concrete, but goal-varying factors were not strongly associated with one another (aside from specificity in Study 1, which was redundant with concreteness and abstractness). Even approach-focus and avoid-focus were not perfectly negatively correlated, nor were concreteness and abstractness. People's New Year's resolutions seem to be mid-level goals, and their content seems to be quite similar across studies, at least in terms of goal domain (Höchli et al., 2019; Norcross & Vangarelli, 1988; Woolley & Fishbach, 2017).

Most people began pursuit with a lot of motivation: high levels of commitment, confidence, and effort. People's motivational trajectories varied, but on average, people had the most motivation at the beginning of the year and less later on. Pursuit behaviors were, like goal content, varied, but could be easily understood as relating to domains like physical health and finances. People varied in the extent to which they told others about their resolutions, and the extent to which they formed habits in their pursuits. Some goal properties are more trait-like, in that people's goals tended to be more similar than not (indicated by high ICCs) and others were more variable within people (indicated by low ICCs). Across studies, ICCs varied, which might be due to differences in measurement. In both studies, Social Commitment and Abstractness were similar across people's resolutions. Most other goal properties were moderate (although Physical Domain was zero in both studies).

There were three common outcomes of goal pursuit across the two studies. One outcome was characterized by success: people reported their status as having achieved their goal (and in some cases, working to maintain it still). Estimates of success varied depending on measurement and were between 20% and 40% across measures in both studies among non-missing observations. Among resolutions that were not achieved, there most were reported as still actively being pursued. Across the studies, between 32% and 60% of resolutions reported as active at the end of the year. The third prevalent outcome was a goal in stasis and "on hold" which was about 15% to 21% of resolutions across studies. Other outcomes of pursuit occurred but were rare. In both studies, very few people deliberately gave up on their goals (between <1% and 3%). About as many people never started (<1% and 3%).

Ordinary goal pursuit seems to operate quite differently than theoretical models assume and imply, and this is most obvious in the observed outcomes of goal pursuit in this study. Many people neither succeeded nor failed, but instead had put their goal on hold. Deliberate disengagement was very uncommon in both studies. Unlike in laboratory settings, ordinary goal pursuit can have a variety of outcomes. Failure to literally achieve goals seems to often relate to the challenge of keeping track of goals and prioritizing one goal among the many other important goals and tasks that people face each day. There is a long history of integrating insights from cognitive psychology into theories of goal pursuit, with particular interest in subtle, non-conscious processes (e.g., priming). However, less subtle effects related to coarse memory and attentional processes seem to play a more important role, as evinced by work on prospective memory (Einstein & McDaniel, 2005), reminders (Vervloet et al., 2012), and implementation intentions (Gollwitzer, 1999). There are factors that operate only outside of the lab – those related to memory, attention, balancing multiple goals, and time – that fundamentally affect the process and outcomes of goal pursuits.

4.1.1 Benefits and challenges of naturalistic, longitudinal survey methodology

The understanding of goal pursuit developed in this dissertation owes to the naturalistic, longitudinal survey methodology taken. This method has many benefits, but it also has challenges. As discussed in the introductory chapter, naturalistic, longitudinal survey research enables ecologically valid work that can support the estimation of effect sizes as they operate in daily life. In addition, in longitudinal work, temporal order is clear, and there is no need to rely on people's memory or perception of their goals and goal pursuits.

However, the naturalistic, longitudinal survey method used here produces messier data relative to experimental research, which can complicate inferences in several ways. For one, missing data due to attrition in longitudinal research is inevitable. Attrition and goal pursuit processes likely have common causes (e.g., personality, skill in self-regulation, as was demonstrated in these data). Thus, attrition may have introduced selection factors that could bias estimates of effects and invalidate inferences. In addition, the observational approach taken here allows factors to covary, but in doing so, confounds the influence of factors. This lack of control (not to mention the lack of experimental manipulation) make clearly reasoning about causality challenging.

One unfulfilled promise of this methodology relates to within and between person effects. I had hoped that the multilevel nature of the survey data in these studies would reveal nuanced dynamics between and within people. However, in both studies but especially Study 2, many people had only one goal. There was, ultimately, insufficient variance within people to meaningfully assess differences of effects between and within people. Future research on resolutions should seek to explore within-person, between-goal effects by collecting data on multiple goals (e.g., at least 3 or 4 per person).

4.2 How well can we predict goal pursuit outcomes?

The various analytic approaches I used throughout this dissertation to explain, and later predict, variance in Subjective Success and Achievement together suggest that most variance in goal pursuit outcomes is unexplained by goal-varying factors and individual differences in self-regulatory skill (and these two explain mostly overlapping variance in goal outcomes). Even powerful machine learning models with dozens of predictors performed barely better than a model that guessed randomly but was informed by the prevalence of success. The best individual predictors – Motivation, Habit Formation, and earlier measures of goal outcomes – don't explain much variance, and also don't offer much useful explanation of the causal process. Most goal-varying factors in this study, and all goal setting factors (i.e., properties of goals like specificity) did not have large, universal, or practically meaningful effects on success.

These studies do not and could not conclusively show that goal-varying factors don't matter universally. However, they do document that goal-varying factors generally don't matter in one reasonable naturalistic setting – New Year's resolutions. This setting is one that is broadly assumed to be meaningfully affected by goal-varying factors (as evidenced by the advice columns published even in reputable news outlets about setting better resolutions every January). The results of this dissertation show that such advice likely unwarranted. Although claims about universal "goal pursuit" effects are common, if taken literally, they are extraordinary. It seems that nothing – not even motivation, likely the most robust factor supported by thousands of empirical studies and hundreds of theories – can sway outcomes of goal pursuit dramatically.

One possibility is that these studies are unique in the relatively weak association that goal-varying properties and skill in self-regulation had with goal outcomes. Several limitations of this study and my modeling approach may have contributed to my null findings. It's possible that measurement error in predictors and outcomes or a failure to consider the right statistical models or the right predictors (or functions of predictors) fully explain the poor performance of the models I estimated.

Another possibility is that goal outcomes – especially when considered across a heterogeneous set of goals – are inherently noisy and subject to stochastic forces. Even people skilled in self-regulation probably don't closely attend to each of their goal pursuits at all times, and likely do not make every choice with total awareness of their goals and consideration for how the choice might affect each of their goals. Other goals, events, and purely chance, arbitrary factors may influence goal outcomes as much or more than factors that theory has identified as causally important to goal pursuit.

4.2.1 Benefits and challenges of estimating and optimizing prediction

The conclusions of this dissertation about prediction were made possible by analytic approaches (e.g., cross-validation) that can estimate and optimize prediction. Theories of self-regulation characterize self-regulatory processes as highly contextualized and dynamic. Yet, our modeling approaches are often coarse, and often describe simple linear relationships. Complex analytic techniques like those I used here are better suited to model self-regulatory processes and can be tailored and highly specific while reducing the risk of overfitting to one set of observations. However, a major challenge of these approaches is that, in addition to being uncommon in psychological science and unfamiliar to many other scholars, they are poorly suited to making inferences. Complex statistical models like SVM may be good at predicting, but they do not provide much information about why and how predictors relate to outcomes.

4.3 Advancing the science of self-regulation

A major limitation of the basic premise of this dissertation - and indeed, of the theories used to develop these studies – is that it brought together very different kinds of goals that likely succeed and fail via different processes and mechanisms. These differences are likely not just of scale, but also of form. For example, smoking cessation requires overcoming physiological barriers in ways that goals like spending more time with family do not. Some goals entail doing a simple behavior often and can easily be automated via habits. Other goals require ongoing adjustment and strategic choice.

Is there such thing as a universal goal pursuit process? It may be that the category of goal pursuit is not a true category. Perhaps it is worthwhile to set aside the goal of trying to explain generic goal processes and instead focus on processes that occur in more narrow contexts, or focus on contexts that give rise to specific and important outcomes, like those related to finances or health. Despite a field-wide tendency to broadly frame research questions and results, most studies of goal pursuit focus on narrow kinds of goals (one off, short-term pursuits), physical health goals, academic goals. The more explicit researchers are about the kinds of goals a model or theory applies to, and the more social psychological research on goal pursuit integrates insights from more focus literatures, the faster our collective understanding of self-regulation and goal pursuit will advance.

Future research should not only seek to be more specific and constrained but should also focus more on description and prediction rather than on elaborating explanatory theoretical models (Yarkoni & Westfall, 2017). Many influential theories in social psychology were rooted in description of ordinary phenomena (Sherif et al., 1954). This basic disposition towards ecological validity and phenomena that occur in daily life can now find new expression in machine learning approaches, which can develop extremely complex, useful data-driven models. Goal research has documented dozens if not hundreds of distinct processes thought to operate in goal pursuit. Those that meaningfully affect goal pursuit in ordinary life are unlikely to have simple, additive effects when operating simultaneously in daily life. No one simple model could explain these processes and their interactions and searching for one by continuing to work on verbal theories rather than mathematical models may not be practically useful for predicting outcomes in the real world.

In addition, for effects that can be instantiated in the laboratory, it will be important to take advantage of the internal validity and inferential tools of the laboratory in order to further test and refine observations from naturalistic contexts (Cialdini, 1980). For example, in this study, several findings suggest that balancing priorities of many goals is a critical barrier to pursuit that leads people to put goals “on hold.” I also found that people rarely deliberately disengage from their goals. Experimental approaches may be useful in parsing people's decision processes and identifying what factors cause people to put goals on hold versus to disengage from them. Critically, any laboratory study of goal pursuit processes must use measures and tasks that have demonstrated reliability and validity and that generalize across stimuli in the laboratory (Clark, 1973; Judd, Westfall, & Kenny, 2012). Establishing that laboratory tasks are valid is not something that is typically rewarded in the way that establishing validity of measures is, but it would nonetheless be an extremely valuable use of research resources that would enable a healthy balance between lab-based and naturalistic research on goal pursuit.

Finally, given how little we know currently about what predicts (and causes) success in ordinary goal pursuit, self-regulation researchers and people reporting on self-regulation research should work to ensure that claims made in research are accompanied by contextual information that can help consumers of research understand, and decide for themselves, whether an effect is practically important. Whether one views an R^2 of .105 as good or bad depends on one's perspective. When interpreted with consideration for the complexity of everyday behavior and the inherent stochasticity of behavior in daily life, explaining 10% of the variance in New Year's resolution outcomes is perhaps impressive. However, when interpreted with consideration of popular claims about the effectiveness of goal-setting strategies, explaining 10% of the variance in New Year's resolution outcomes is less impressive. Goal-varying properties that are often described to the public as causally related to better pursuit outcomes increased the predictive accuracy of an empty model by just 3%, which amounts to about 30 resolutions in Study 2. Reasonable people would disagree about whether the effect sizes found in this study (and in other studies of goal pursuit) matter, and whether they are large enough to warrant public-facing advice about goal pursuit. Consumers of research would likely have varying opinions, too, if given full context. Communicating effect size information and acknowledging small effects or large portions of unexplained variance can empower consumers of research to make judgements about practical significance for themselves.

Appendix A: Supplemental Results for Study 1

Table 34: Correlations of Each Goal-Varying Property with Success Derived from the Entire Dataset in Study 1

Variable	*r*	CI
Physical Domain	0.01	-.07, .09
Mental Domain	-0.02	-.10, .06
Specificity	0.05	-.03, .13
Concrete-Abstract	0.04	-.04, .11
Avoid-Approach	-0.04	-.12, .04
Motivation	0.07	-.01, .15
Habit Formation	.20**	.12, .27
Social Commitment	.24**	.16, .32

Note. Subjective Success $M = 2.98$, $SD = 1.42$, $N = 1094$

Table 35: Means, Standard Deviations, Ns, and Correlations of Goal-Varying Properties Derived from the Entire Dataset in Study 1

Variable	*M*	*SD*	*N*	1	2	3	4	5	6	7	8	9
1. Physical Domain	1.56	0.50	854									
2. Mental Domain	1.49	0.50	854	.06								
3. Specific-Relative	0.47	0.50	1090	.03	-.01							
4. Specific-Vague	0.14	0.34	1090	.02	.17**	-.37**						
5. Concrete-Abstract	1.07	0.26	1088	-.12**	.20**	-.19**	.60**					
6. Avoid-Approach	0.89	0.31	1046	-.19**	.04	.11**	.02	-.04				
7. Motivation	4.19	0.82	1094	-.05	-.07*	-.08**	.02	-.02	.10**			
8. Habit Formation	13.62	6.11	716	.05	-.07	.03	.01	-.02	.04	.30**		
9. Social Commitment	2.90	1.55	808	.03	-.12**	.01	-.06	-.08*	.02	.18**	.14**	
10. Trait Self-Control	3.51	0.78	1094	-.03	-.09**	-.03	-.07*	-.06*	.04	.23**	.11**	.13**

Note. M and *SD* are used to represent mean and standard deviation, respectively. *N* is used to indicate the available sample size for each variable. Variables derived from the survey administered in January (T1) were Specificity, Concrete-Abstract, Avoid-Approach, Motivation, and Trait Self-Control. Dummy coded variables list the level coded as zero first and the level coded as one second. Specific-Relative and Specific-Vague are dummy coded Specificity variables. * indicates $p < .05$. ** indicates $p < .01$.

Table 36: Bayesian Multilevel Model Regressing Physical Domain on Subjective Success in Study 1

Predictors	*Estimates*	*CI*
(Intercept)	2.94	2.75,3.12
Physical	0.06	-0.19,0.29
Trait Self-Control	0.29	0.14,0.44
Random Effects		
σ^2	1.35	1.26,1.45
τ_{00}	0.42	0.07,0.71
τ_{11}	0.24	0.01,0.61
ρ_{01}	-0.34	-0.98,0.86
N	235	
Observations	604	
Marginal R^2 / Conditional R^2	0.028 / 0.111	

Note. CI represents credible interval. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts.

Table 37: Bayesian Multilevel Model Regressing Mental Domain on Subjective Success in Study 1

Predictors	*Estimates*	*CI*
(Intercept)	2.96	2.79,3.12
Mental	0.02	-0.21,0.24
Trait Self-Control	0.29	0.14,0.44
Random Effects		
σ^2		
τ_{00}	0.35	0.04,0.64
τ_{11}	0.31	0.01,0.77
ρ_{01}	-0.13	-0.94,0.91
N	235	
Observations	604	
Marginal R^2 / Conditional R^2	0.028 / 0.107	

Note. CI represents credible interval. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts.

Table 38: Bayesian Multilevel Model Regressing Concrete-Abstract on Subjective Success in Study 1

Predictors	*Estimates*	*CI (95%)*
Intercept	2.97	2.85,3.09
Concrete-Abstract	-0.16	-0.64,0.30
Trait Self-Control	0.26	0.11,0.41
Random Effects		
σ^2	1.36	1.27,1.45
τ_{00}	0.34	0.04,0.57
N	248	
Observations	631	
Marginal R^2 / Conditional R^2	0.024 / 0.088	

Note. CI represents credible interval. Concrete was coded as zero, abstract was coded as 1. σ^2 represents residual variance. τ_{00} represents random intercept variance.

Table 39: Bayesian Multilevel Model Regressing Specificity on Subjective Success in Study 1

Predictors	*Estimates*	*CI (95%)*
Intercept	2.83	2.65,3.00
Specific-Non-Specific	0.22	-0.02,0.45
Specific-Vague	0.30	-0.06,0.66
Trait Self-Control	0.28	0.14,0.42
Random Effects		
σ^2	1.36	1.27,1.45
τ_{00}	0.33	0.05,0.57
N	248	
Observations	632	
Marginal R^2 / Conditional R^2	0.033 / 0.091	

Note. CI represents credible interval. Specific was coded as zero for both dummy variables. σ^2 represents residual variance. τ_{00} represents random intercept variance.

Table 40: Bayesian Multilevel Model Regressing Avoid-Approach on Subjective Success in Study 1

Predictors	*Estimates*	*CI (95%)*
Intercept	2.68	2.34,3.01
Avoid-Approach	0.30	-0.05,0.65
Trait Self-Control	0.27	0.11,0.42
Random Effects		
σ^2	1.35	1.25,1.44
τ_{00}	0.36	0.02,0.82
τ_{11}	0.31	0.01,0.80
ρ_{01}	-0.26	-0.97,0.90
N	244	
Observations	604	
Marginal R^2 / Conditional R^2	0.030 / 0.106	

Note. CI represents credible interval. Avoid was coded as zero, approach was coded as 1. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts.

Table 41: Bayesian Multilevel Model Regressing Motivation on Subjective Success in Study 1

Predictors	*Estimates*	*CI*
(Intercept)	2.97	2.86,3.08
Motivation (goal)	0.26	0.06,0.46
Motivation (person)	0.35	0.15,0.55
Trait Self-Control	0.18	0.03,0.33
Random Effects		
σ^2	1.34	1.25,1.43
τ_{00}	0.31	0.02,0.55
τ_{11}	0.27	0.02,0.62
ρ_{01}	-0.13	-0.96,0.93
N	250	
Observations	636	
Marginal R^2 / Conditional R^2	0.051 / 0.128	

Note. CI represents credible interval. Motivation (goal) is the group-mean centered Level 1 predictor. Motivation (person) is the grand-mean centered Level 2 predictor. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts.

Table 42: Bayesian Multilevel Model Regressing Social Commitment on Subjective Success in Study 1

Predictors	*Estimates*	*CI*
Intercept	2.95	2.82,3.07
Social Commitment (goal)	0.05	-0.06,0.17
Social Commitment (person)	0.09	-0.01,0.19
Trait Self-Control	0.26	0.10,0.42
Random Effects		
σ^2	1.33	1.24,1.43
τ_{00}	0.38	0.08,0.61
τ_{11}	0.10	0.00,0.26
ρ_{01}	-0.18	-0.96,0.91
N	219	
Observations	566	
Marginal R^2 / Conditional R^2	0.038 / 0.123	

Note. Results from the Bayesian analysis. CI represents credible interval. Social Commitment (goal) is the group-mean centered Level 1 predictor. Social Commitment (person) is the grand-mean centered Level 2 predictor. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts.

Appendix B: Supplemental Results for Study 2

Table 43: Correlations of Each Goal-Varying Property with Success Derived from the Entire Sample in Study 2

Variable	*r*	CI
1. Physical Domain	0.02	-.05, .10
2. Mental Domain	-0.02	-.09, .05
3. Concrete	0.04	-.04, .11
4. Abstract	-0.03	-.10, .04
5. Approach-Focus	0.02	-.05, .09
6. Avoidance-Focus	-0.01	-.08, .06
7. Motivation	.21**	.14, .28
8. Habit Formation	.32**	.24, .39
9. Social Commitment	.08*	.01, .15

Table 44: Means, Standard Deviations, N, and Correlations of Goal-Varying Properties Derived from the Entire Sample in Study 2

Variable	*M*	*SD*	*N*	1	2	3	4	5	6	7	8	9
1. Physical Domain	1.40	0.49	1342									
2. Mental Domain	1.48	0.50	1342	-.06*								
3. Concrete	4.24	1.06	1335	-.07*	-.06*							
4. Abstract	2.60	1.52	1335	.06*	.10**	-.55**						
5. Approach	4.18	1.18	1340	.12**	.07*	.07*	.04					
6. Avoid	3.39	1.62	1338	-.30**	.00	.08**	.11**	-.29**				
7. Motivation	4.37	0.66	1342	.04	.03	.25**	-.02	.13**	.05			
8. Habit Formation	3.55	2.07	766	.07*	.05	.01	.08*	.12**	-.02	.24**		
9. Social Commitment	3.46	1.53	1340	-.03	-.01	.19**	-.01	.05	.07**	.28**	.09**	
10. Trait Self-Control	3.55	0.73	1342	-.07*	-.01	.10**	-.05	.08**	-.05	.29**	.15**	.13**

Note. *M* and *SD* are used to represent mean and standard deviation, respectively, and values are not accurate sample estimates because they were derived from the entire sample of resolutions (i.e., people with multiple resolutions were overrepresented). *N* is used to indicate the available sample size for each variable. Habit Formation was derived from the survey administered in July (T2). All other variables were derived from the survey administered in January (T1). Dummy coded variables list the level coded as zero first and the level coded as one second. * indicates $p < .05$. ** indicates $p < .01$.

Table 45: Means, Standard Deviations, and N of Additional January (T1) Predictors in the Entire Dataset in Study 2

Variable	*M*	*SD*	*N*	*Min*	*Max*
Society Domain*	0.06	0.23	1342	0	1
Spiritual Domain*	0.08	0.28	1342	0	1
Social Domain*	0.85	0.36	1342	0	1
Money Domain*	0.31	0.46	1342	0	1
Family Domain*	0.83	0.38	1342	0	1
Career Domain*	0.14	0.35	1342	0	1
Education Domain*	0.07	0.25	1342	0	1
Other Domain*	0.02	0.15	1342	0	1
New Resolution*	0.40	0.49	1342	0	1
Substance Use*	0.09	0.28	1360	0	1
Self-Control Required	4.40	0.89	1337	1	5
Cost	2.77	1.44	1339	1	5
Value	4.73	0.55	1338	1	5
Track Progress	4.10	1.17	1337	1	5
Plan How	4.47	0.78	1338	1	5
Plan When	4.27	1.00	1340	1	5
Anticipate Barriers	4.23	0.98	1339	1	5
Percent Progress	30.69	28.38	1352	0	100
Age	36.70	11.85	1324	18	76

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. * indicates a binary variable. Binary variables list the level coded as zero first and the level coded as one second.

Table 46: Means, Standard Deviations, and N of Additional July (T2) Predictors in the Entire Dataset in Study 2

Variable	*M*	*SD*	*N*	*Min*	*Max*
Motivation	3.72	1.17	774	1	5
Achievement*	0.42	0.49	774	0	1
Deliberately Stopped	1.61	0.49	187	1	5
Forgot Resolution*	0.34	0.47	774	0	1
Modified Resolution*	1.86	0.35	774	0	1
Made Easier-Harder	2.71	1.21	108	1	5
Status- Missing*	0.42	0.50	1360	0	1
Status- On Hold*	0.09	0.29	1360	0	1
Status- Achieved*	0.10	0.30	1360	0	1
Percent Progress	44.79	32..59	769	1	100

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. * indicates a binary variable. Binary variables list the level coded as zero first and the level coded as one second.

Table 47: Frequentist and Bayesian Multilevel Models Regressing Physical Domain on Subjective Success in Study 2

	Frequentist		*Bayesian*	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.87 ***	2.75,2.99	2.87	2.75,2.99
Physical	0.08	-0.11,0.26	0.08	-0.11,0.27
Trait Self-Control	0.17 *	0.04,0.30	0.17	0.04,0.30
Random Effects				
σ^2	1.41		1.19	1.10,1.29
τ_{00}	0.23		0.41	0.01,0.64
τ_{11}			0.31	0.01,0.73
ρ_{01}			0.09	-0.84,0.95
N_{people}	449			
$N_{resolutions}$	749			
Marginal R^2 / Conditional R^2	0.010 / 0.147		0.011 / 0.149	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. Physical was coded as one if present and zero if absent. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. **p* <0.05 ** *p* <0.01 ****p* <0.001

Table 48: Frequentist and Bayesian Multilevel Models Regressing Physical Domain on Subjective Success in Study 2

	Frequentist		***Bayesian***	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.87 ***	2.75,2.99	2.87	2.75,2.99
Physical	0.08	-0.11,0.26	0.08	-0.11,0.27
Trait Self-Control	0.17 *	0.04,0.30	0.17	0.04,0.30
Random Effects				
σ^2	1.41		1.19	1.10,1.29
τ_{00}	0.23		0.41	0.01,0.64
τ_{11}			0.31	0.01,0.73
ρ_{01}			0.09	-0.84,0.95
N_{people}	449			
$N_{resolutions}$	749			
Marginal R^2 / Conditional R^2	0.010 / 0.147		0.011 / 0.149	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. Physical health domain was coded as one if present and zero if absent. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. *p <0.05 ** p <0.01 ***p <0.001

Table 49: Frequentist and Bayesian Multilevel Models Regressing Mental Domain on Subjective Success in Study 2

	Frequentist		*Bayesian*	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.92 ***	2.79,3.05	2.92	2.79,3.05
Mental	-0.04	-0.23,0.14	-0.04	-0.24,0.16
Trait Self-Control	0.17 *	0.03,0.30	0.16	0.03,0.29
Random Effects				
σ^2	1.42		1.20	1.11,1.30
τ_{00}	0.23		0.40	0.04,0.64
τ_{11}			0.32	0.02,0.81
ρ_{01}			-0.10	-0.93,0.92
N_{people}	449			
$N_{resolutions}$	749			
Marginal R^2 / Conditional R^2	0.009 / 0.145		0.011 / 0.139	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. Mental was coded as one if present and zero if absent. σ^2 represents residual variance. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. **p* <0.05 ** *p* <0.01 ****p* <0.001

Table 50: Frequentist and Bayesian Multilevel Models Regressing Goal Concreteness on Subjective Success in Study 2

	Frequentist		*Bayesian*	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.90 ***	2.80,3.00	2.90	2.80,3.00
Concrete (goal)	0.09	-0.06,0.23	0.08	-0.07,0.24
Concrete (person)	0.01	-0.10,0.11	0.01	-0.11,0.12
Trait Self-Control	0.17 *	0.03,0.31	0.17	0.03,0.31
Random Effects				
σ^2	1.40		1.19	1.10,1.29
τ_{00}	0.24		0.47	0.18,0.66
τ_{11}			0.12	0.00,0.33
ρ_{01}			-0.07	-0.95,0.92
N_{people}	445			
$N_{resolutions}$	741			
Marginal R^2 / Conditional R^2	0.011 / 0.156		0.014 / 0.156	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. **p* <0.05 ** *p* <0.01 ****p* <0.001

Table 51: Frequentist and Bayesian Multilevel Models Regressing Goal Abstractness on Subjective Success in Study 2

	Frequentist		*Bayesian*	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.90 ***	2.80,2.99	2.90	2.79,2.99
Abstract (goal)	-0.01	-0.10,0.08	-0.08	-0.20,0.04
Abstract (person)	0.09	-0.06,0.23	-0.00	-0.08,0.07
Trait Self-Control	0.17 *	0.04,0.30	0.16	0.03,0.29
Random Effects				
σ^2	1.41		1.20	1.11,1.30
τ_{00}	0.23		0.44	0.14, 0.63
τ_{11}			0.09	0.00,0.25
ρ_{01}			0.12	-0.92,0.96
N_{people}	447			
$N_{resolutions}$	745			
Marginal R^2 / Conditional R^2	0.011 / 0.149		0.013 / 0.137	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. *p <0.05 ** p <0.01 ***p <0.001

Table 52: Frequentist and Bayesian Multilevel Models Regressing Avoid-Focus on Subjective Success in Study 2

	Frequentist		***Bayesian***	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.89 ***	2.80,2.99	2.89	2.80,3.00
Avoid (goal)	-0.03	-0.12,0.06	-0.02	-0.12,0.08
Avoid (person)	-0.01	-0.08,0.07	-0.01	-0.08,0.07
Trait Self-Control	0.17 *	0.04,0.30	0.17	0.04,0.30
Random Effects				
σ^2	1.42		1.19	1.10,1.29
τ_{00}	0.23		0.47	0.21,0.66
τ_{11}			0.11	0.00,0.27
ρ_{01}			-0.23	-0.96,0.86
N_{people}	445			
$N_{resolutions}$	742			
Marginal R^2 / Conditional R^2	0.010 / 0.148		0.013 / 0.157	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. Avoid (goal) is the group-mean centered Level 1 predictor. Avoid (person) is the grand-mean centered Level 2 predictor. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. *p <0.05 ** p <0.01 ***p <0.001

Table 53: Frequentist and Bayesian Multilevel Models Regressing Approach-Focus on Subjective Success in Study 2

	Frequentist		***Bayesian***	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.90 ***	2.80,2.99	2.90	2.80,2.99
Approach (goal)	-0.01	-0.10,0.08	0.09	-0.07,0.24
Approach (person)	0.09	-0.06,0.23	-0.01	-0.10,0.08
Trait Self-Control	0.17 *	0.04,0.30	0.17	0.03,0.30
Random Effects				
σ^2	1.41		1.20	1.11,1.29
τ_{00}	0.23		0.46	0.19,0.65
τ_{11}			0.13	0.00,0.36
ρ_{01}			-0.02	-0.95,0.94
N_{people}	447			
$N_{resolutions}$	745			
Marginal R^2 / Conditional R^2	0.011 / 0.149		0.014 / 0.149	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. Approach (goal) is the group-mean centered Level 1 predictor. Approach (person) is the grand-mean centered Level 2 predictor. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. *p <0.05 ** p <0.01 ***p <0.001

Table 54: Frequentist and Bayesian Multilevel Models Regressing Motivation (T1) on Subjective Success in Study 2

	Frequentist		***Bayesian***	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.91 ***	2.81,3.01	2.91	2.81,3.01
Motivation (goal)	0.46 ***	0.19,0.73	0.48	0.19,0.77
Motivation (person)	0.37 ***	0.20,0.55	0.37	0.20,0.55
Trait Self-Control	0.07	-0.07,0.21	0.07	-0.07,0.21
Random Effects				
σ^2	1.37		1.17	1.07,1.27
τ_{00}	0.21		0.45	0.16,0.65
τ_{11}			0.38	0.02,0.91
ρ_{01}			0.15	-0.88,0.94
N_{people}	449			
$N_{resolutions}$	749			
Marginal R^2 / Conditional R^2	0.046 / 0.172		0.050 / 0.188	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. Motivation (goal) is the group-mean centered Level 1 predictor. Motivation (person) is the grand-mean centered Level 2 predictor. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. **p* <0.05 ** *p* <0.01 ****p* <0.001

Table 55: Frequentist and Bayesian Multilevel Models Regressing Social Commitment on Subjective Success in Study 2

	Frequentist		***Bayesian***	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.91 ***	2.81,3.00	2.91	2.81,3.01
Social Commitment (goal)	0.01	-0.12,0.13	0.00	-0.13,0.13
Social Commitment (person)	0.08 *	0.00,0.15	0.08	0.00,0.15
Trait Self-Control	0.15 *	0.01,0.28	0.15	0.01,0.29
Random Effects				
σ^2	1.42		1.20	1.11,1.30
τ_{00}	0.22		0.44	0.13,0.64
τ_{11}			0.13	0.01,0.32
ρ_{01}			0.25	-0.86,0.97
N_{people}	447			
$N_{resolutions}$	746			
Marginal R^2 / Conditional R^2	0.016 / 0.147		0.018 / 0.149	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. Social Commitment (goal) is the group-mean centered Level 1 predictor. Social Commitment (person) is the grand-mean centered Level 2 predictor. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. *p <0.05 ** p <0.01 ***p <0.001

Table 56: Frequentist and Bayesian Multilevel Models Regressing Habit Formation on Subjective Success in Study 2

	Frequentist		***Bayesian***	
Predictors	*Estimate*	*CI*	*Estimate*	*CI*
(Intercept)	2.92 ***	2.82,3.03	2.92	2.82,3.03
Habit Formation (goal)	0.20 ***	0.12,0.28	0.21	0.12,0.30
Habit Formation (person)	0.20 ***	0.13,0.26	0.20	0.13,0.26
Trait Self-Control	0.09	-0.05,0.24	0.10	-0.05,0.25
Random Effects				
σ^2	1.36		1.14	1.02,1.25
τ_{00}	0.16		0.43	0.11,0.65
τ_{11}			0.19	0.02,0.34
ρ_{01}			-0.20	-0.93,0.81
N_{people}	340			
$N_{resolutions}$	581			
Marginal R^2 / Conditional R^2	0.105 / 0.198		0.110 / 0.251	

Note. CI represents confidence interval for the frequentist results and credible interval for Bayesian results. Habit Formation (goal) is the group-mean centered Level 1 predictor. Habit Formation (person) is the grand-mean centered Level 2 predictor. τ_{00} represents random intercept variance. τ_{11} represents random slope variance. ρ_{01} represents the correlation between random slopes and intercepts. **p* <0.05 ** *p* <0.01 ****p* <0.001

References

Albarracín, D., Wilson, K., Chan, M. S., Durantini, M., & Sanchez, F. (2018). Action and inaction in multi-behaviour recommendations: A meta-analysis of lifestyle interventions. *Health Psychology Review*, *12*(1), 1–24. https://doi.org/10.1080/17437199.2017.1369140

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, *64*(6), 359–372.

Austin, J. T., & Vancouver, J. B. (1996). Goal constructs in psychology: Structure, process, and content. *Psychological Bulletin*, *120*(3), 338–375. https://doi.org/10.1037/0033-2909.120.3.338

Avishai, A., Conner, M., & Sheeran, P. (2019). Setting realistic health goals: Antecedents and consequences. *Annals of Behavioral Medicine*, *53*(12), 1020–1031. https://doi.org/10.1093/abm/kaz012

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122–147. https://doi.org/10.1037/0003-066X.37.2.122

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Berkman, E. T., Hutcherson, C. A., Livingston, J. L., Kahn, L. E., & Inzlicht, M. (2017). Self-control as value-based choice. *Current Directions in Psychological Science*, *26*(5), 422–428. https://doi.org/10.1177/0963721417704394

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, 21(2), 205–229. https://doi.org/10.1177/0049124192021002004

Bosworth, H. B., Blalock, D. V., Hoyle, R. H., Czajkowski, S. M., & Voils, C. I. (2018). The role of psychological science in efforts to improve cardiovascular medication adherence. *American Psychologist*, *73*(8), 968–980. https://doi.org/10.1037/amp0000316

Brandstatter, V., Herrmann, M., & Schuler, J. (2013). The struggle of giving up personal goals: Affective, physiological, and cognitive consequences of an action crisis.

Personality and Social Psychology Bulletin, *39*(12), 1668–1682. https://doi.org/10.1177/0146167213500151

Brandstätter, V., & Schüler, J. (2013). Action crisis and cost–benefit thinking: A cognitive analysis of a goal-disengagement phase. *Journal of Experimental Social Psychology*. http://www.sciencedirect.com/science/article/pii/S0022103112002065

Brodscholl, J. C., Kober, H., & Higgins, E. T. (2007). Strategies of self-regulation in goal attainment versus goal maintenance: Attainment and maintenance. *European Journal of Social Psychology*, *37*(4), 628–648. https://doi.org/10.1002/ejsp.380

Brown, J. M., Miller, W. R., & Lawendowski, L. a. (1999). The Self-Regulation Questionnaire (SRQ). *Innovations in Clinical Practice: A Source Book (Vol. 17)*, 281–289.

Brunstein, J. C. (1993). Personal goals and subjective well-being: A longitudinal study. *Journal of Personality and Social Psychology*, *65*(5), 1061–1070. https://doi.org/10.1037/0022-3514.65.5.1061

Bryan, C. J., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences*, *108*(31), 12653–12656. https://doi.org/10.1073/pnas.1103343108

Bupa (2015). *New Year's Resolutions.* https://www.comresglobal.com/wp-content/uploads/2015/12/BUPA_NY-Resolution_Public-Polling_Nov-15_UPDATED-TABLES.pdf

Bürkner, P.-C. (2017). **brms**: An *R* Package for Bayesian Multilevel Models Using *Stan*. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Carden, L., & Wood, W. (2018). Habit formation and change. *Current Opinion in Behavioral Sciences*, *20*, 117–122. https://doi.org/10.1016/j.cobeha.2017.12.009

Carver, C. S., & Scheier, M. F. (1982). Control theory: A useful conceptual framework for personality–social, clinical, and health psychology. *Psychological Bulletin*, *92*(1), 111–135. https://doi.org/10.1037/0033-2909.92.1.111

Carver, C., & Scheier, M. (2011). Self-regulation of action and affect. In K. D. Vohs & R. F. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (2nd ed., pp. 3–21). Guilford Press.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of verbal learning and verbal behavior*, *12*(4), 335-359.

Chen, P.-H. A. (2016). *Can Self-Regulation Be Improved? An Examination of Neural and Behavioral Evidence: Vol. ProQuest Number 10250544*.

Chen, P.-H. A., Kelley, W. M., Lopez, R. B., & Heatherton, T. F. (2019). Reducing reward responsivity and daily food desires in female dieters through domain-specific training. *Social Neuroscience*, *14*(4), 470–483. https://doi.org/10.1080/17470919.2018.1495667

Converse, B. A., Juarez, L., & Hennecke, M. (2019). Self-control and the reasons behind our goals. *Journal of Personality and Social Psychology*, *116*(5), 860–883. https://doi.org/10.1037/pspp0000188

Dai, H., Milkman, K. L., & Riis, J. (2014). The fresh start effect: temporal landmarks motivate aspirational behavior. *Management Science*, *60*(10), 2563–2582. https://doi.org/10.1287/mnsc.2014.1901

Danaei, G., Ding, E. L., Mozaffarian, D., Taylor, B., Rehm, J., Murray, C., & Ezzati, M. (2009). The preventable causes of death in the United States: Comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLOS Medicine*, *6*(4).

Darnon, C., Harackiewicz, J. M., Butera, F., Mugny, G., & Quiamzade, A. (2007). Performance-approach and performance-avoidance goals: When uncertainty makes a difference. *Personality & Social Psychology Bulletin*, *33*(6), 813–827. https://doi.org/10.1177/0146167207301022

Davis, S. W., & Haws, K. L. (2017). Don’t sweat the big stuff: emphasizing importance hinders goal pursuit for consumers low in dispositional self-control resources. *Journal of the Association for Consumer Research*, *2*(1), 93–104. https://doi.org/10.1086/690115

Davydenko, M., Werner, K. M., Milyavskaya, M., Donnellan, M. B., & Donnellan, M. B. (2019). Frozen goals: identifying and defining a new type of goal. *Collabra: Psychology*, *5*(1), 17. https://doi.org/10.1525/collabra.194

Deci, E. L., & Ryan, R. M. (2000). The " what " and " why " of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 37–41. https://doi.org/10.1207/S15327965PLI1104

Ding, Y., Tang, S., Liao, S. G., Jia, J., Oesterreich, S., Lin, Y., & Tseng, G. C. (2014). Bias correction for selecting the minimal-error classifier from many machine learning models. *Bioinformatics*, *30*(22), 3152-3158.

Di Paula, A., & Campbell, J. D. (2002). Self-esteem and persistence in the face of failure. *Journal of Personality and Social Psychology*, *83*(3), 711–724.

Duckworth, A. L., Gendler, T. S., & Gross, J. J. (2016). Situational strategies for self-control. *Perspectives on Psychological Science*, *11*(1), 35–55. https://doi.org/10.1177/1745691615623247

Duckworth, Angela L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology*, *14*(1), 91–118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Ebersole, C. R., Atherton, O. E., Belanger, A. L., Skulborstad, H. M., Allen, J. M., Banks, J. B., Baranski, E., Bernstein, M. J., Bonfiglio, D. B. V., Boucher, L., Brown, E. R., Budiman, N. I., Cairo, A. H., Capaldi, C. A., Chartier, C. R., Chung, J. M., Cicero, D. C., Coleman, J. A., Conway, J. G., … Nosek, B. A. (2016). Many Labs 3: Evaluating participant pool quality across the academic semester via replication. *Journal of Experimental Social Psychology*, *67*, 68–82. https://doi.org/10.1016/j.jesp.2015.10.012

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Einstein, G. O., & McDaniel, M. A. (2005). Prospective memory: Multiple retrieval processes. *Current Directions in Psychological Science*, *14*(6), 286–290. https://doi.org/10.1111/j.0963-7214.2005.00382.x

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*(1), 2319. https://doi.org/10.1038/s41467-019-10301-1

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, *70*(3), 461–475. https://doi.org/10.1037/0022-3514.70.3.461

Elliot, A. J., & Thrash, T. M. (2002). Approach-avoidance motivation in personality: Approach and avoidance temperaments and goals. *Journal of Personality and Social Psychology*, *82*(5), 804–818. https://doi.org/10.1037/0022-3514.82.5.804

Elliot, A. J., & Fryer, James. (2008). The goal construct in psychology. In *Handbook of motivation science*, (pp. 235–250).

Emmons, R. A. (1986). Personal strivings: An approach to personality and subjective well-being. *Journal of Personality and Social Psychology*, *51*(5), 1058–1068. https://doi.org/10.1037/0022-3514.51.5.1058

Emmons, R. A., & Diener, E. (1986). A goal-affect analysis of everyday situational choices. *Journal of Research in Personality*, *20*(3), 309–326. https://doi.org/10.1016/0092-6566(86)90137-6

Enders, C. K. (2017). Multiple imputation as a flexible tool for missing data handling in clinical research. *Behaviour Research and Therapy*, *98*, 4–18. https://doi.org/10.1016/j.brat.2016.11.008

Etkin, J. (2019). Time in relation to goals. *Current Opinion in Psychology*, *26*, 32–36. https://doi.org/10.1016/j.copsyc.2018.04.013

Fishbach, A., & Finkelstein, S. (2012). How feedback influences persistence, disengagement, and change in goal pursuit. In H. Aarts & A. Elliot (Eds.), *Goal-directed behavior* (pp. 203–230). Psychology Press.

Fitzsimons, G. M., Finkel, E. J., & Vandellen, M. R. (2015). Transactive goal dynamics. *Psychological Review*, *122*(4), 648–673.

Flake, J. K., Barron, K. E., Hulleman, C., McCoach, B. D., & Welsh, M. E. (2015). Measuring cost: The forgotten component of expectancy-value theory. *Contemporary Educational Psychology*, *41*, 232–244. https://doi.org/10.1016/j.cedpsych.2015.03.002

Freund, A. M., & Hennecke, M. (2015). On means and ends: The role of goal focus in successful goal pursuit. *Current Directions in Psychological Science*, *24*(2), 149–153. https://doi.org/10.1177/0963721414559774

Friedman, H. S., Tucker, J. S., Schwartz, J. E., Martin, L. R., Tomlinson-Keasey, C., Wingard, D. L., & Criqui, M. H. (1995). Childhood conscientiousness and longevity: Health behaviors and cause of death. *Journal of Personality and Social Psychology*, *68*(4), 696–703. https://doi.org/10.1037/0022-3514.68.4.696

Fujita, K. (2011). On conceptualizing self-control as more than the effortful inhibition of impulses. *Personality and Social Psychology Review*, *15*(4), 352–366. https://doi.org/10.1177/1088868311411165

Fujita, K., Trope, Y., Liberman, N., & Levin-Sagi, M. (2006). Construal levels and self-control. *Journal of Personality and Social Psychology*, *90*(3), 351–367. https://doi.org/10.1037/0022-3514.90.3.351

Galla, B. M., & Duckworth, A. L. (2015). More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. *Journal of Personality and Social Psychology*, *109*(3), 508–525. https://doi.org/10.1037/pspp0000026

Gardner, B. (2015). A review and analysis of the use of 'habit' in understanding, predicting and influencing health-related behaviour. *Health Psychology Review*, *9*(3), 277–295. https://doi.org/10.1080/17437199.2013.876238

Gardner, B., Abraham, C., Lally, P., & de Bruijn, G.-J. (2012). Towards parsimony in habit measurement: Testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. *International Journal of Behavioral Nutrition and Physical Activity*, *9*(1), 102. https://doi.org/10.1186/1479-5868-9-102

Ghassemi, M., Bernecker, K., Herrmann, M., & Brandstätter, V. (2017). The process of disengagement from personal goals: reciprocal influences between the experience of action crisis and appraisals of goal desirability and attainability. *Personality and Social Psychology Bulletin*, *43*(4), 524-537.

Gollwitzer, P. M. (1999). Implementation intentions: Strong effects of simple plans. *American Psychologist*, *54*(7), 493–503. https://doi.org/10.1037/0003-066X.54.7.493

Goodman, F. R., Kashdan, T. B., Stiksma, M. C., & Blalock, D. V. (2019). Personal strivings to understand anxiety disorders: Social anxiety as an exemplar. *Clinical Psychological Science, 7*(2), 283–301. https://doi.org/10.1177/2167702618804778

Gravert, C. A., & Olsson Collentine, L. (2019). When nudges aren't enough: incentives and habit formation in public transport usage. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3500699

Guan, J., Xiang, P., & Keating, X. D. (2004). Evaluating the replicability of sample results: A tutorial of double cross-validation methods. *Measurement in physical education and exercise science*, *8*(4), 227-241.

Heath, C., Larrick, R. P., & Wu, G. (1999). Goals as reference points. *Cognitive Psychology, 38*(1), 79–109. https://doi.org/10.1006/cogp.1998.0708

Hennecke, M. (2019). What doesn't kill you . . .: Means for avoidance goal pursuit are less enjoyable than means for approach goal pursuit. *Motivation Science, 5*(1), 1–13. https://doi.org/10.1037/mot0000104

Hennecke, M., & Converse, B. A. (2017). Next week, next month, next year: how perceived temporal boundaries affect initiation expectations. *Social Psychological and Personality Science*, *8*(8), 918–926. https://doi.org/10.1177/1948550617691099

Hennecke, M., Czikmantori, T., & Brandstätter, V. (2018). Doing despite disliking: Self-regulatory strategies in everyday aversive activities. *European Journal of Personality*, per.2182. https://doi.org/10.1002/per.2182

Herrmann, M., & Brandstätter, V. (2015). Action crises and goal disengagement: Longitudinal evidence on the predictive validity of a motivational phase in goal striving. *Motivation Science*, *1*(2), 121–136. https://doi.org/10.1037/mot0000016

Höchli, B., Brügger, A., & Messner, C. (2018). How focusing on superordinate goals motivates broad, long-term goal pursuit: A theoretical perspective. *Frontiers in Psychology*, *9*, 1879. https://doi.org/10.3389/fpsyg.2018.01879

Höchli, B., Brügger, A., & Messner, C. (2019). Making new year's resolutions that stick: Exploring how superordinate and subordinate goals motivate goal pursuit. *Applied Psychology: Health and Well-Being*, aphw.12172. https://doi.org/10.1111/aphw.12172

Hofmann, W., Baumeister, R. F., Förster, G., & Vohs, K. D. (2012). Everyday temptations: An experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology*, *102*(6), 1318–1335. https://doi.org/10.1037/a0026545

Hollenbeck, J. R., & Klein, H. J. (1987). Goal commitment and the goal-setting process: Problems, prospects, and proposals for future research. *Journal of Applied Psychology*, *72*(2), 212–220. https://doi.org/10.1037/0021-9010.72.2.212

Houben, K., Nederkoorn, C., Wiers, R. W., & Jansen, A. (2011). Resisting temptation: Decreasing alcohol-related affect and drinking behavior by training response inhibition. *Drug and Alcohol Dependence*, *116*(1–3), 132–136. https://doi.org/10.1016/j.drugalcdep.2010.12.011

Hoyle, R. H., & Davisson, E. K. (2016). Varieties of self-control and their personality correlates. In *Handbook of self-regulation: Research, theory, and applications*.

Hoyle, R. H., & Davisson, E. K. (2018). Self-Regulation: An Integrative Review. In V. Zeigler-Hill & T. Shackelford, *The SAGE Handbook of Personality and Individual Differences: Volume III: Applications of Personality and Individual Differences* (pp. 115–131). SAGE Publications Ltd. https://doi.org/10.4135/9781526451248.n5

Hoyle, R. H., & Gallagher, P. (2015). *The Interplay of Personality and Self-Regulation*. *4*, 189–207.

Jackson, J. J., Wood, D., Bogg, T., Walton, K. E., Harms, P. D., & Roberts, B. W. (2010). What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness (BIC). *Journal of Research in Personality*, *44*(4), 501–511. https://doi.org/10.1016/j.jrp.2010.06.005

James, G., Witten, D., Hastie, T., & Tibshirani, R. (Eds.). (2013). *An introduction to statistical learning: With applications in R*. Springer.

Ji, M., & Wood, W. (2007). Purchase and Consumption Habits: Not Necessarily What You Intend. *Journal of Consumer Psychology*, *17*(4), 261–276. https://doi.org/10.1016/S1057-7408(07)70037-2

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524–532. https://doi.org/10.1177/0956797611430953

John, O. P., Donahue, E. M., & L, Kentle. R. (1991). *The Big Five Inventory—Versions 4a and 54* (Vol. 37). https://doi.org/10.1016/S0092-6566(03)00046-1

Jonathan, P., Krzanowski, W. J., & McCarthy, W. V. (2000). On the use of cross-validation to assess performance in multivariate prediction. *Statistics and Computing*, *10*(3), 209-229.

Joireman, J., & Kuhlman, D. M. (2004). The Zuckerman-Kuhlman Personality Questionnaire: Origin, development, and validity of a measure to assess an alternative five-factor model of personality. In *On the Psychobiology of Personality* (pp. 49–64). Elsevier. https://doi.org/10.1016/B978-008044209-9/50005-1

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of personality and social psychology*, *103*(1), 54.

King, L. A., & Broyles, S. J. (1997). Wishes, gender, personality, and well-being. *Journal of Personality*, *65*(1), 49–76. https://doi.org/10.1111/j.1467-6494.1997.tb00529.x

King, L. A., Richards, J. H., & Stemmerich, E. (1998). Daily goals, life goals, and worst fears: Means, ends, and subjective well-being. *Journal of Personality*, *66*(5), 713–744. https://doi.org/10.1111/1467-6494.00030

Klinger, E. (1975). Consequences of commitment to and disengagement from incentives. *Psychological Review*, *82*(1), 1–25. https://doi.org/10.1037/h0076171

Koo, M., & Fishbach, A. (2008). Dynamics of self-regulation: How (un)accomplished goal actions affect motivation. *Journal of Personality and Social Psychology*, *94*(2), 183–195. https://doi.org/10.1037/0022-3514.94.2.183

Koo, M., & Fishbach, A. (2012). The small-area hypothesis: Effects of progress monitoring on goal adherence. *Journal of Consumer Research*, *39*(3), 493–509. https://doi.org/10.1086/663827

Köpetz, C., Faber, T., Fishbach, A., & Kruglanski, A. W. (2011). The multifinality constraints effect: How goal multiplicity narrows the means set to a focal end. *Journal of Personality and Social Psychology*, *100*(5), 810–826. https://doi.org/10.1037/a0022980

Kristal, A. S., & Whillans, A. V. (2020). What we can learn from five naturalistic field experiments that failed to shift commuter behaviour. *Nature Human Behaviour*, *4*(2), 169–176. https://doi.org/10.1038/s41562-019-0795-z

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Woo Young, C., & Sleeth-Keppler, D. (2002). A theory of goal systems. *Advances in Experimental Social Psychology, Volume 34*, 331–378. https://doi.org/10.1016/S0065-2601(02)80008-9

Kuhn, M. (2008). Building Predictive Models in *R* Using the **caret** Package. *Journal of Statistical Software*, *28*(5). https://doi.org/10.18637/jss.v028.i05

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Lally, P., van Jaarsveld, C. H. M., Potts, H. W. W., & Wardle, J. (2010). How are habits formed: Modelling habit formation in the real world. *European Journal of Social Psychology*, *40*(6), 998–1009. https://doi.org/10.1002/ejsp.674

Latham, G. P., & Locke, E. a. (1991). Self-regulation through goal setting. *Organizational Behavior and Human Decision Processes*, *50*(2), 212–247. https://doi.org/10.1016/0749-5978(91)90021-K

Little, B. R., Lecci, L., & Watkinson, B. (1992). Personality and personal projects: Linking Big Five and PAC units of analysis. *Journal of Personality*, *60*(2), 501–525.

Locke, E. A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, *3*(2), 157–189. https://doi.org/10.1016/0030-5073(68)90004-4

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969—1980. *Psychological Bulletin*, *90*(1), 125–152.

Ludwig, R. M., Srivastava, S., Berkman, E. T., & Donnellan, B. (2018). Planfulness: A process-focused construct of individual differences in goal achievement. *Collabra: Psychology*, *4*(1). https://doi.org/10.1525/collabra.136

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. *Science*, *341*, 976–980. https://doi.org/10.1126/science.1238041

McAdams, D. P. (1996). Personality, modernity, and the storied self: A contemporary framework for studying persons. *Psychological Inquiry*, *7*(4), 295–321. https://doi.org/10.1207/s15327965pli0704_1

McAuliffe, W. H., Moshontz, H., McCauley, T. G., & McCullough, M. E. (2020). Searching for prosociality in qualitative data: Comparing manual, closed-vocabulary, and open-vocabulary methods. *European Journal of Personality*. https://doi.org/10.1002/per.2240 https://doi.org/10.31234/osf.io/y3fj4

McFarlin, D. B., Baumeister, R. F., & Blascovich, J. (1984). On knowing when to quit: Task failure, self-esteem, advice, and nonproductive persistence. *Journal of Personality*, *52*(2), 138–155. https://doi.org/10.1111/j.1467-6494.1984.tb00349.x

McGuire, J. T., & Kable, J. W. (2013). Rational temporal predictions can underlie apparent failures to delay gratification. *Psychological Review*, *120*(2), 395–410. https://doi.org/10.1037/a0031910

Medicine, N. R. C. and I. of. (2015). *Measuring the Risks and Causes of Premature Death: Summary of Workshops*.

Micha, R., Penalvo, J., Cudhea, F., Imamura, F., Rehm, C., & Mozaffarian, D. (2017). Association between dietary factors and mortality from heart disease, stroke, and type 2 diabetes in the United States. *JAMA*, *317*(9), 912–924.

Miller, D. I. (2019). When do growth mindset interventions work? *Trends in Cognitive Sciences*, *23*(11), 910–912. https://doi.org/10.1016/j.tics.2019.08.005

Milyavskaya, M., Berkman, E. T., & De Ridder, D. T. D. (2018). The many faces of self-control: Tacit assumptions and recommendations to deal with them. *Motivation Science*. https://doi.org/10.1037/mot0000108

Milyavskaya, M., & Inzlicht, M. (2017). What's so great about self-control? Examining the importance of effortful self-control and temptation in predicting real-life depletion and goal attainment. *Social Psychological and Personality Science*, *8*(6), 603–611. https://doi.org/10.1177/1948550616679237

Milyavskaya, M., & Nadolny, D. (2018). Health goals among American adults: Prevalence, characteristics, and barriers. *Journal of Health Psychology*, *23*(10), 1350–1355. https://doi.org/10.1177/1359105316656227

Neal, Wood, W., & Drolet, A. (2013). How do people adhere to goals when willpower is low? The profits (and pitfalls) of strong habits. *Journal of Personality and Social Psychology*, *104*(6), 959–975. https://doi.org/10.1037/a0032626

Norcross, J. C., & Vangarelli, D. J. (1988). The resolution solution: Longitudinal examination of New Year's change attempts. *Journal of Substance Abuse*, *1*(2), 127–134. https://doi.org/10.1016/S0899-3289(88)80016-6

Oishi, S., & Diener, E. (2001). Goals, culture, and subjective well-being. *Personality and Social Psychology Bulletin*, *27*(12), 1674–1682. https://doi.org/10.1177/01461672012712010

Orben, A., & Przybylski, A. K. (2019). The association between adolescent well-being and digital technology use. *Nature Human Behaviour*, *3*(2), 173.

Oreopoulos, P., & Petronijevic, U. (2019). *The remarkable unresponsiveness of college students to nudging and what we can learn from it* (No. w26059; p. w26059). National Bureau of Economic Research. https://doi.org/10.3386/w26059

Reisz, Z., Boudreaux, M. J., & Ozer, D. J. (2013). Personality traits and the prediction of personal goals. *Personality and Individual Differences*, *55*(6), 699–704. https://doi.org/10.1016/j.paid.2013.05.023

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment*, *31*(12), 1395–1411. https://doi.org/10.1037/pas0000754

Roberts, B. W., O'Donnell, M., & Robins, R. W. (2004). Goal and personality trait development in emerging adulthood. *Journal of Personality and Social Psychology*, *87*(4), 541–550. https://doi.org/10.1037/0022-3514.87.4.541

Robson, D. A., Allen, M. S., & Howard, S. J. (2020). Self-regulation in childhood as a predictor of future outcomes: A meta-analytic review. *Psychological Bulletin*. https://doi.org/10.1037/bul0000227

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Rogers, T., Milkman, K. L., & Volpp, K. G. (2014). Commitment devices: Using initiatives to change behavior. *JAMA*, *311*(20), 2065. https://doi.org/10.1001/jama.2014.3485

Rohrer, J. (2018). The uncanny mountain: P-values between .01 and .10 are still a problem. *The 100% CI*. http://www.the100.ci/2018/02/15/the-uncanny-mountain-p-values-between-01-and-10-are-still-a-problem/

Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2017). Probing birth-order effects on narrow traits using specification-curve analysis. *Psychological Science*, *28*(12), 1821–1832. https://doi.org/10.1177/0956797617723726

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Salmela-Aro, K., Read, S., Nurmi, J.-E., Vuoksimaa, E., Siltala, M., Dick, D. M., Pulkkinen, L., Kaprio, J., & Rose, R. J. (2012). Personal goals and personality traits among young adults: Genetic and environmental effects. *Journal of Research in Personality*, *46*(3), 248–257. https://doi.org/10.1016/j.jrp.2012.01.007

Saunders, B., Milyavskaya, M., Etz, A., Randles, D., & Inzlicht, M. (2018). Reported self-control is not meaningfully associated with inhibition-related executive function: A bayesian analysis. *Collabra: Psychology*, *4*(1), 39. https://doi.org/10.1525/collabra.134

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. https://doi.org/10.1037/1082-989X.7.2.147

Schimmack, U. (2012). The ironic effect of significant results on the credibility of multiple-study articles. *Psychological Methods*, *17*(4), 551–566. https://doi.org/10.1037/a0029487

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). *False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything*. https://doi.org/10.1177/0956797611417632

Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., & Stine-Morrow, E. A. L. (2016). Do “brain-training” programs work? *Psychological Science in the Public Interest*, *17*(3), 103–186. https://doi.org/10.1177/1529100616661983

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*(2), 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Simmons, J. P. & Nelson, L. D. (2015). Specification curve: Descriptive and inferential statistics on all reasonable specifications. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2694998

Stautz, K., Zupan, Z., Field, M., & Marteau, T. M. (2018). Does self-control modify the impact of interventions to change alcohol, tobacco, and food consumption? A systematic review. *Health Psychology Review*, *12*(2), 157–178. https://doi.org/10.1080/17437199.2017.1421477

Steegen, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702–712. https://doi.org/10.1177/1745691616658637

Stillman, P. E., Medvedev, D., & Ferguson, M. J. (2017). Resisting temptation: Tracking how self-control conflicts are successfully resolved in real time. *Psychological Science*, *28*(9), 1240–1258. https://doi.org/10.1177/0956797617705386

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*(2), 271–324. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*(2), 440–463. https://doi.org/10.1037/a0018963

Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*(1), 3–15. https://doi.org/10.1037/0033-295X.94.1.3

Veilleux, J. C., Hill, M. A., Skinner, K. D., Pollert, G. A., Spero, K. D., & Baker, D. E. (2018). Self-control failure scenarios in daily life: Developing a taxonomy of goals and temptations. *Motivation and Emotion*, *42*(5), 653–670. https://doi.org/10.1007/s11031-018-9695-1

Vervloet, M., Linn, A. J., van Weert, J. C. M., de Bakker, D. H., Bouvy, M. L., & van Dijk, L. (2012). The effectiveness of interventions using electronic reminders to improve adherence to chronic medication: A systematic review of the literature. *Journal of the American Medical Informatics Association*, *19*(5), 696–704. https://doi.org/10.1136/amiajnl-2011-000748

Wallace, S. G., & Etkin, J. (2018). How goal specificity shapes motivation: a reference points perspective. *Journal of Consumer Research*, *44*(5), 1033–1051. https://doi.org/10.1093/jcr/ucx082

Weigel, D., & Murray, C. (2000). The paradox of stability and change in relationships: What does chaos theory offer for the study of romantic relationships? *Journal of*

Social and Personal Relationships, *17*(3), 425–449. https://doi.org/10.1177/0265407500173006

Weinberg, R., Bruya, L., & Jackson, A. (1985). The effects of goal proximity and goal specificity on endurance performance. *Journal of Sport Psychology*, *7*(3), 296–305. https://doi.org/10.1123/jsp.7.3.296

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wilkowski, B. M., & Ferguson, E. L. (2016). The steps that can take us miles: Examining the short-term dynamics of long-term daily goal pursuit. *Journal of Experimental Psychology: General*, *145*(4), 516–529. https://doi.org/10.1037/xge0000150

Wood, W., Quinn, J. M., & Kashy, D. A. (2002). Habits in everyday life: Thought, emotion, and action. *Journal of Personality and Social Psychology*, *83*(6), 1281–1297. https://doi.org/10.1037/0022-3514.83.6.1281

Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, *67*(1), 289–314. https://doi.org/10.1146/annurev-psych-122414-033417

Woolley, K., & Fishbach, A. (2016). For the fun of it: harnessing immediate rewards to increase persistence in long-term goals. *Journal of Consumer Research*, *42*(6), 952–966. https://doi.org/10.1093/jcr/ucv098

Wright, P. M., & Kacmar, K. M. (1994). Goal specificity as a determinant of goal commitment and goal change. *Organizational Behavior and Human Decision Processes*, *59*(2), 242–260. https://doi.org/10.1006/obhd.1994.1059

Wrosch, C., Scheier, M. F., & Miller, G. E. (2013). Goal adjustment capacities, subjective well-being, and physical health. *Social and Personality Psychology Compass*, *7*(12), 847–860. https://doi.org/10.1111/spc3.12074.

Wrosch, C., Scheier, M. F., Miller, G. E., Schulz, R., & Carver, C. S. (2003). Adaptive self-regulation of unattainable goals: Goal disengagement, goal reengagement, and subjective well-being. *Personality and Social Psychology Bulletin*, *29*(12), 1494–1508. https://doi.org/10.1177/0146167203256921

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yeager, D. S., Hanselman, P., Walton, G. M., Murray, J. S., Crosnoe, R., Muller, C., Tipton, E., Schneider, B., Hulleman, C. S., Hinojosa, C. P., Paunesku, D., Romero, C., Flint, K., Roberts, A., Trott, J., Iachan, R., Buontempo, J., Yang, S. M., Carvalho, C. M., … Dweck, C. S. (2019). A national experiment reveals where a growth mindset improves achievement. *Nature*, *573*(7774), 364–369. https://doi.org/10.1038/s41586-019-1466-y

YouGov (2017). New Year's Resolutions. https://d25d2506sfb94s.cloudfront.net/cumulus_uploads/document/2041y00vbb/InternalResults_170106_NYResolutions_W.pdf

YouGov (2019). New Year's Resolutions. https://d25d2506sfb94s.cloudfront.net/cumulus_uploads/document/307bcd90p8/Results%20for%20Editorial%20(NYRs2019)%20293%207.12.2018.xlsx%20%20%5bGroup%5d.pdf

Zuckerman, M. (2008). Zuckerman-Kuhlman Personality Questionnaire (ZKPQ): An operational definition of the alternative five factorial model of personality. In *The SAGE Handbook of Personality Theory and Assessment: Volume 2—Personality Measurement and Testing* (pp. 219–238). SAGE Publications Ltd. https://doi.org/10.4135/9781849200479.n11

Biography

Hannah Moshontz de la Rocha received a B.A. in Psychology from Reed College in 2011 and an M.A. in Psychology and Neuroscience from Duke University in 2016.

She has authored and co-authored nine published and in-press articles and book chapters titled: Reporting standards for literature searches and report inclusion: Making research syntheses more transparent and easy to replicate; Self-grading and Peer-grading for formative and summative assessments in 3rd Through 12th grade classrooms: A meta-analysis; Who makes the grade? A synthesis of research comparing self, peer and instructor grades in college classrooms; The Psychological Science Accelerator: Advancing psychology through a distributed collaborative network; Seven easy steps to open science: An annotated reading list; A meta-analysis on the impact of grades and comments on academic motivation and achievement: A case for written feedback; Self-regulation: An individual difference perspective; Self-control and affect regulation styles predict anxiety longitudinally in university students; Uncovering prosociality in qualitative data: Comparing manual, closed-vocabulary, and open-vocabulary methods.

Since obtaining her bachelor's degree, she has received the following honors and fellowships: Summer Research Fellowship for First- and Second-Year Ph.D. Students, Duke University; Education Human Development Scholar Fellowship, Duke Social Science Research Institute; Graduate Summer Research Fellowship, Duke University; and the Program for Advanced Research in the Social Sciences Fellowship, Duke University.