Mathematical and Computational Modeling of Suicide as a Complex Dynamical System

Shirley B. Wang, A.M.[1,3]

Donald J. Robinaugh, Ph.D.[2,3]

Alexander J. Millner, Ph.D.[1]

Rebecca G. Fortgang, Ph.D. [1,3]

Matthew K. Nock, Ph.D.[1,3]

[1]Department of Psychology, Harvard University, Cambridge, MA

[2]Department of Applied Psychology, Northeastern University, Boston, MA

[3]Department of Psychiatry, Massachusetts General Hospital, Boston, MA

Correspondence should be addressed to Shirley B. Wang, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, MA 02138. Email: shirleywang@g.harvard.edu

Acknowledgements

Shirley B. Wang is supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1745303 and the National Institute of Mental Health (NIMH) NRSA Predoctoral Fellowship F31MH125495. Donald J. Robinaugh is supported by NIMH Career Development Award K23MH113805. Alexander J. Millner is supported by NIMH R34MH124973. Rebecca G. Fortgang is supported by NIMH Career Development Award K23MH132766. Matthew K. Nock is supported by NIMH U01MH116928. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Mental Health or the National Science Foundation.

Disclosures

Dr. Nock receives publication royalties from Macmillan, Pearson, and UpToDate. He has been a paid consultant in the past three years for Microsoft Corporation, the Veterans Health Administration, and COMPASS Pathways, and for legal cases regarding a death by suicide. He has stock options in Cerebral Inc. He is an unpaid scientific advisor for Empatica, Koko, and TalkLife.

Abstract

Background: Despite decades of research, the current suicide rate is nearly identical to what it was 100 years ago. This slow progress is due, at least in part, to a lack of formal theories of suicide. Existing suicide theories are instantiated *verbally,* omitting details required for precise explanation and prediction, rendering them difficult to effectively evaluate and difficult to improve. By contrast, formal theories are instantiated *mathematically* and *computationally,* allowing researchers to precisely deduce theory predictions, rigorously evaluate what the theory can and cannot explain, and thereby, inform how the theory can be improved. This paper takes the first step toward addressing the need for formal theories in suicide research by formalizing an initial, general theory of suicide and evaluating its ability to explain suicide-related phenomena.

Methods: First, we formalized a General Escape Theory of Suicide as a system of stochastic and ordinary differential equations. Second, we used these equations to simulate behavior of the system over time. Third, we evaluated if the formal theory produced robust suicide-related phenomena including rapid onset and brief duration of suicidal thoughts, and zero-inflation of suicidal thinking in time series data.

Results: Simulations successfully produced the proposed suicidal phenomena (i.e., rapid onset, short duration, and high zero-inflation of suicidal thoughts in time series data). Notably, these simulations also produced theorized phenomena following from the General Escape Theory of Suicide: that suicidal thoughts emerge when alternative escape behaviors failed to effectively regulate aversive internal states, and that effective use of long-term strategies may prevent the emergence of suicidal thoughts.

Conclusions: To our knowledge, the model developed here is the first formal theory of suicide, which was able to produce – and, thus, explain – well-established phenomena documented in the suicide literature. We discuss the next steps in a research program dedicated to studying suicide as a complex dynamical system, and describe how the integration of formal theories and empirical research may advance our understanding, prediction, and prevention of suicide.

***Keywords*:** suicide; mathematical modeling; formal theory; computational psychiatry

Suicide is a leading cause of death worldwide, claiming over 800,000 lives each year (WHO, 2018). Over 47,000 people die by suicide each year in the United States alone, making it the 12th leading cause of death overall, the second leading cause of death for young people ages 10 – 34, and among the most devastating of all public health problems (CDC, 2021). Despite a near-exponential increase in suicide research over the past several decades (Franklin et al., 2017), our collective progress in understanding, predicting, and preventing suicide has been too slow. Until recently, the prediction of suicidal thoughts and behaviors has remained just slightly above chance for all outcomes, with no meaningful improvement in predictive accuracy across 50 years of research from 1965 – 2015 (Franklin et al., 2017). Intervention effect sizes have also remained small, with no single intervention for suicidal thoughts and behaviors appearing consistently stronger than others, and no meaningful improvement in intervention efficacy over time (Fox et al., 2020). Whereas scientific advances have led to significant declines in other leading causes of death over time (e.g., pneumonia, tuberculosis), the suicide rate today is nearly identical to what it was 100 years ago (CDC, 2021; Fortgang & Nock, 2021).

This slow progress is due (at least in part) to the methods we typically use in clinical psychology to develop, express, and evaluate our theories (Borsboom et al., 2021; Muthukrishna & Henrich, 2019; van Rooij & Baggio, 2020). Psychologists have theorized about suicide for decades, but these have all been *verbal theories*: theories expressed in natural language. Due to the inherent imprecision of language, verbal theories are often underspecified and contain hidden assumptions or contradictions (Epstein, 2008; Fried, 2020; Guest & Martin, 2020; Millner et al., 2020; Smaldino, 2017). For instance, many theories suggest that suicide is an escape from aversive psychological states (e.g., aversive self-awareness, psychological pain) (Baumeister, 1990; Linehan, 1987; Shneidman, 1993). However, like most verbal theories, they do not specify the strength, form, or time scale of theorized effects. This leaves fundamental questions unanswered: for example, at what level of aversive psychological states do these theories predict will result in suicide as an attempt to escape? What is the shape of this relationship (do increases in aversive states cause linear or nonlinear increases in suicidal thoughts and behaviors)? And crucially, if someone is experiencing a requisite level of aversive psychological states, *when* should suicidal thoughts or behaviors emerge (the same day, the next day, the next week)?

Verbal theories typically do not provide this level of detail, and the resultant ambiguities make it difficult (if not impossible) to understand precisely what they are proposing to explain or predict. Thus, verbal theories are limited to making vague predictions (e.g., "negative affect causes suicidal thoughts") that tend to rely on *evaluation via null-hypothesis significance testing* (e.g., finding a correlation between negative affect and suicidal thinking). Such statistical significance tests are easy to pass because nearly all constructs in psychopathology research are intercorrelated to some extent (i.e., the "crud factor") (Meehl, 1990a). This makes verbal theories difficult to falsify, because even a small correlation among variables would be seen as providing support to the theory – and also difficult to convincingly corroborate, as any observed correlation would provide minimal evidence in favor of the theory (Millner et al., 2020). As a result, it is challenging to determine the extent to which verbal theories accord with empirical data, and how they can be used to guide clinical intervention efforts (Epstein, 2008; Robinaugh et al., 2021).

Formal theories overcome many of these limitations by expressing theories in the language of mathematics. Through analytic deduction or computer simulations, formal theories allow researchers to directly observe how a system of interest will behavior according to a given theory. This allows for evaluation of a theory *based on its ability to reproduce real-world phenomena*. Notably, the historical reliance on verbal theories in psychopathology research stands in stark contrast to other sciences that devote considerable resources and effort to formal theory development (e.g., physics, biology, ecology). The theoretical branches of these sciences have yielded important insights. For instance, formalizing theories in ecology revealed how ecosystems (e.g., lakes, forests) experience sudden catastrophic shifts and collapses, allowing theoretical biologists to identify early warning signs of critical shifts in ecosystems, and identify targets to improve their resilience (Scheffer et al., 2001, 2009; Van Nes & Scheffer, 2007).

Formal theories may be especially useful for studying suicide as a *complex dynamical system*. By complex dynamical system, we mean a system composed of many components (e.g., emotions, cognitions, behaviors, environmental factors) characterized by nonlinear interactions, feedback loops, stochasticity, and dynamic change over time (Boccara, 2010; Fisher & Pruitt, 2020; Olthof et al., 2023; Ottino, 2003; Siegenfeld & Bar-Yam, 2020). Here, we also distinguish between *complicated* and *complex* systems. The behavior of complicated systems results from a series of clear cause-and-effect relationships, meaning that a complicated system can be understood by breaking it down into its constituent parts and studying each component (and its specific interactions with other components) individually. However, this is not the case for complex systems, which are characterized by *emergent phenomena*: behavior that emerges from the complex interactions among components (with no central controller). To understand the behavior of a complex system, we must study the system as a whole, rather than examining each component in isolation (i.e., *"the whole is greater than the sum of its parts"*). Therefore, while systems such as machines can be studied as complicated systems, systems in *life* – from ecosystems, to social systems, to psychological systems – are ideally studied as complex.

Among the myriad mathematical modeling methods available, systems of differential equations are especially useful for studying change complex systems over time. Differential equation models are used across the natural and physical sciences to formalize theories, from population models (e.g., Lotka-Volterra models) to infectious disease spread (e.g., Susceptible-Infectious-Recovered models) to electrodynamics (e.g., Maxwell's equations). Applying these methods to study mental disorders as complex systems, Robinaugh and colleagues (2019) recently built a formal model of panic disorder as a system of differential equations. Simulations from this model successfully produced several known phenomena about panic attacks (e.g., rapid onset of arousal and perceived threat out of the blue, efficacy of cognitive behavioral therapy). Just as importantly, this model also identified limitations in existing verbal theories. For instance, in early iterations of its development, the model was not able to produce the robust phenomena of non-clinical panic attacks. In model simulations, every "person" who experienced panic attacks eventually developed panic disorder, illustrating the need for ongoing theory refinement to account for non-clinical panic attacks. The ability of formal models to reveal blind spots and unknowns – and to allow for incremental improvements over time – is one of the most important advantages of formal theories relative to verbal theories. We believe that such an approach could also be critical for advancing the understanding of suicide, and for identifying key topics for future research (Millner et al., 2020).

In this paper, we take a first step towards bringing formal modeling into suicide research. We formalize several core components of a General Escape Theory of Suicide, a new theory of suicidal thoughts and behaviors that our team has recently developed (Millner et al., in prep) which brings together common factors and processes proposed in long-standing theories of suicide (e.g., Baumeister, 1990; Joiner, 2005; Klonsky & May, 2015; Shneidman, 1993) into a single overarching framework and seeks to ground those processes in basic psychological science. Accordingly, the model illustrated here (described in more detail in Millner et al., in prep) builds on the foundation provided by earlier verbal theories of suicide by formalizing core components of those theories as a mathematical model. Importantly, we do not believe that the General Escape Theory is a conclusive (final) theory of suicide, nor is our goal to propose a final formalization in this paper. Instead, we propose that this initial mathematical formalization may equips us to better evaluate the theory, identify its strengths and shortcomings, and determine the empirical research and areas for further theory development that will be needed if we are to make genuine advancements in our understanding of suicide and our ability to prevent it.

This paper is organized as follows. First, in Section 1, we introduce (verbally) the core components of the General Escape Theory of Suicide (see **Figure 1**), specifically focusing on components that give rise to suicidal thoughts. In Section 2, we describe our formalization of the theory as a mathematical model, and implement these equations computationally to simulate artificial theory-implied data, allowing for direct observation of dynamics predicted by the theory over time. Using these simulations, we examine if the formal theory can successfully produce (1) robust suicide-related phenomena, and (2) core theoretical predictions following from the General Escape Theory of Suicide. Finally, in Section 3, we discuss the implications of this initial formalization effort for the ongoing development of theories of suicide.

Section 1: Core Components of a General Escape Theory of Suicide

The central premise of our General Escape Theory of Suicide (Millner et al., in prep) is that suicidal thoughts and behaviors arise from a desire to escape aversive internal states, and function to reduce those aversive internal states. That is, stressors experienced through the course of day-to-day life produce aversive internal states (e.g., negative affect, psychological pain), which lead to an urge to escape (Baumeister, 1990; Gee & Casey, 2015; McLaughlin & Hatzenbuehler, 2009). This, in turn, leads people to engage in *escape behaviors* oriented towards achieving immediate and short-term relief from distress. Importantly, we take a *functional* approach to characterize escape behaviors. That is, we define escape behaviors by the processes that maintain and reinforce them, rather than by their topographical characteristics (Kazdin, 2012; Nock & Prinstein, 2004; Wang et al., 2021). Therefore, at this stage, the theory does not make a distinction between canonically "adaptive" forms of escape behaviors (e.g., going for a run) and canonically "maladaptive" escape behaviors (e.g., disordered eating, substance use, nonsuicidal self-injury). Instead, we focus on the *purpose* of these behaviors: to immediately escape from aversive internal states without effectively tolerating or regulating them, or addressing their proximal cause.

Escape behaviors are likely – at least initially – to successfully reduce aversive internal states. In the language of dynamical systems, there is a dampening feedback loop between the aversive internal state and escape behavior. As the aversive internal state becomes more intense, the urge to escape grows, and more escape behavior is engaged in with the intent of regulating the aversive internal state. In this way, escape behaviors become reinforced because they help manage the experience of aversive internal states. Although escape behaviors serve as effective regulators, at least for a time, they also have limitations: they do not help people effectively tolerate or cope with aversive internal states, nor do they address the stressors that elicited distress, thereby failing to reduce the likelihood of experiencing aversive internal states in the longer-term. Escape behaviors are, therefore, a short-term strategy.

By contrast, *long-term strategies* require individuals to tolerate aversive internal states without immediately seeking escape. The General Escape Theory of Suicide distinguishes between two forms of long-term strategies: *internal-focused* strategies (which intervene on aversive internal states, e.g., cognitive reappraisal), and *external-focused* strategies (which intervene on external stressors, e.g., problem-solving). For example, if someone experiences distress because they lost their job, external-focused strategies could include starting a new job search or updating one's resume, and internal-focused strategies could include practicing radical acceptance or nonjudgmental awareness. Of note, the key distinction between internal-focused strategies and escape behaviors is that internal-focused strategies do not emerge from the urge to escape aversive internal states, but rather require *tolerance* of the discomfort to effectively regulate distress (in other words, the urge to escape has a positive effect on escape behaviors, but a negative effect on internal-focused strategies) (Lass et al., 2020; Peterson et al., 2014; Wang & Borders, 2018). Accordingly, in the context of aversive internal states, individuals can use escape behaviors (to rapidly diminish distress) or slower, long-term strategies that require tolerance of the initial discomfort but may lead to sustained reductions in distress over time, through intervening on the aversive internal state or the stressors driving it. Importantly, given our focus on *functional* definitions of theory constructs, the same observed behavior (e.g., exercise) could be identified as different theory components (e.g., escape behavior, external-focused strategy, internal-focused strategy) for different people (even within individuals over time), leading to different outcomes depending on the context and purpose of the behavior.

This General Escape Theory of Suicide posits that suicidal thoughts and behaviors arise from the urge to escape, and aim to eliminate aversive internal states (Baumeister, 1990; Bentley et al., 2021; Gunn III, 2014; Kleiman et al., 2018; Kuehn et al., 2022). Suicidal thoughts and behaviors can thus be understood as a specific type of escape behavior. We propose that suicidal thoughts emerge when one's short- and long-term strategies fail to reduce aversive internal states and, in turn, the urge to escape. The reason for this failure may vary, but one common path may be that escape behaviors dimmish in effectiveness through repeated use. Clinically, this reflects *tolerance*, or the diminished effect of escape behaviors (e.g., nonsuicidal self-injury, alcohol use, disordered eating) with repeated exposure. When this diminished efficacy occurs in the context of low tolerance for the aversive internal state – itself another consequence of repeated reliance on escape behaviors – then aversive internal states will be experienced as persistent, severe, uncontrollable, and inescapable. This leads individuals to turn toward higher-cost escape behaviors in order to escape their aversive internal states, including, ultimately, thinking about suicide. Like other escape strategies, suicidal thoughts reduce aversive internal states in the short-term, reinforcing those thoughts and making the individual more likely to turn to them in the future as a means of regulating overwhelming emotions (Hennings, 2020; Kleiman et al., 2018; Kuehn et al., 2022). Of note, while the complete General Escape Theory of Suicide (Millner et al., in prep) also describes pathways from suicidal thoughts to suicidal behaviors, we focused our initial theory formalization efforts on modeling the dynamics of a system including suicidal thoughts only, as shown in **Figure 1**.

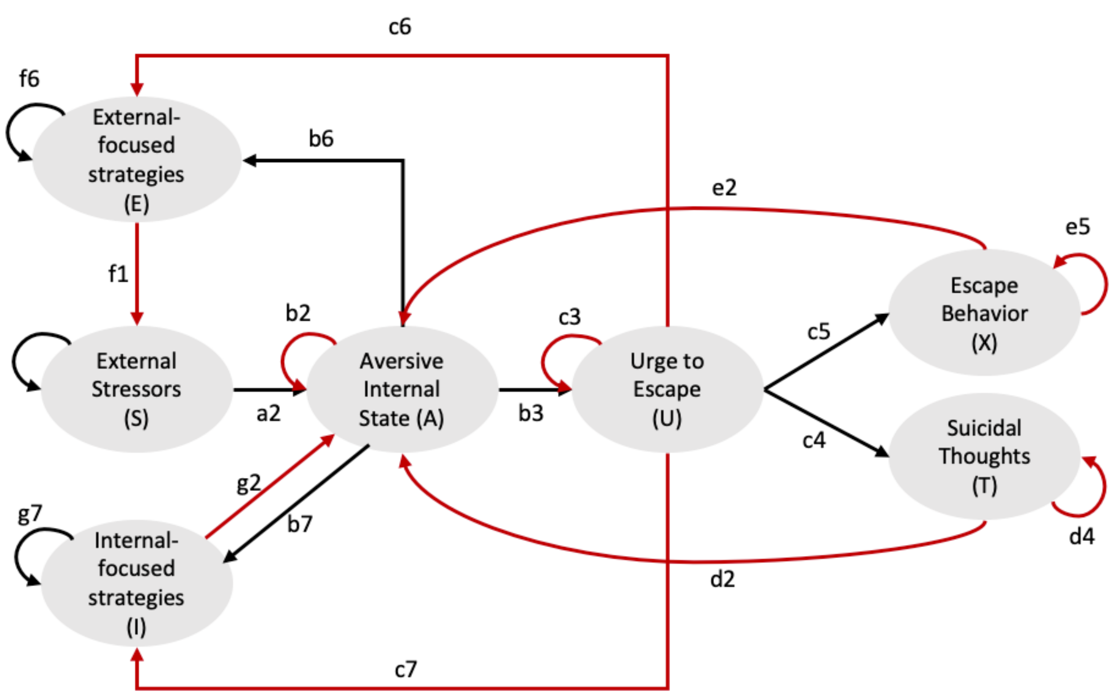

Figure 1. Core components of the General Escape Theory of Suicide. Black arrows represent positive effects and red arrows represent negative effects.

Section 2: Formalizing the General Escape Theory of Suicide as a Mathematical Model

In this section, we formalize core components of the General Escape Theory of Suicide (**Figure 1**) as a system of differential equations, implement these equations as a computational model, and simulate from the model to evaluate its predictions. All code for implementing the model is publicly available at github.com/ShirleyBWang/math_model_suicide. To make each model component and its simulated behavior as clear as possible, we build up the model one step at a time, starting with stressors only, and then adding in aversive internal states, urge to escape, suicidal thoughts, escape behaviors, and finally internal- and external-focused strategies.

At each step, we first present the *mathematical model* formalizing that theory component. Second, we implement the equations to simulate two weeks of model behavior. Of note, we formalized all components as ordinary differential equations (ODEs), except for stressors, which we formalized as a *stochastic* differential equation (SDE). While ODEs are deterministic, SDEs are stochastic and include a random noise component. Thus, we simulated multiple *realizations* (simulations with the same parameter values that generate varying dynamics due to stochasticity) to observe a distribution of potential dynamics over time. Third, we examined if simulations produced three robust suicide-related phenomena: (a) relatively rapid onset and short duration of suicidal thoughts, (b) high variability and dynamic fluctuations in suicidal thoughts, and (c) zero-inflation in time series data of suicidal thoughts (Czyz et al., 2022; Kivelä et al., 2022; Kleiman et al., 2017; Nock et al., 2009; Sedano-Capdevila et al., 2021; Wang et al., 2021; Wang et al., 2023). Finally, we manipulated model parameters to evaluate the model's ability to produce key phenomena from the General Escape Theory of Suicide, including (a) the emergence of suicidal thoughts when escape behaviors do not effectively regulate aversive internal states, and that (b) engaging in effective long-term strategies may prevent the emergence of suicidal thoughts.

This section is intended to be as accessible as possible to readers without a background in this type of mathematical and computational modeling. Although a complete introduction to differential equations is beyond the scope of this paper (for good introductions, we refer interested readers to Campbell & Haberman, 2011; Farlow, 2006; Mongin et al., 2022; Ricardo, 2020), we provide a short overview of these models here. Briefly, differential equations are a family of mathematical equations that relate one or more functions and their derivatives, such as $\frac{dx}{dt} = 2x + 4y$. The left-hand side of an ODE (the type of differential equation we primarily use in this paper) represents the derivative with respect to one independent variable (typically with respect to time), such as $\frac{dx}{dt}$. In this case, if x represents suicidal thoughts, $\frac{dx}{dt}$ represents the *instantaneous rate of change* in suicidal thoughts over time. The right-hand side of an ODE describes the function: the processes that defines how the variable changes over time, such as $2x + 4y$. In this case, the equation specifies that change in x over time is determined by x itself and another variable y. The *solution* to an ODE is found by integrating the equation. In this case:

$$\int \frac{dx}{dt} = \int 2x + 4y \tag{1.1}$$

$$\int \frac{1}{2x + 4y} dx = \int dt \tag{1.2}$$

$$\frac{1}{2} ln(2x + 4y) = t + C \tag{1.3}$$

$$2x + 4y = Ce^{2t} \tag{1.4}$$

$$x(t) = Ce^{2t} - 2y \tag{1.5}$$

Thus, while an ODE describes how a variable changes over time, its *solution* describes the state of that variable (in this case, the value of x) at a given time. Of note, many ODEs used in real-life applications do not have closed form solutions (i.e., there is no explicit mathematical formula that solves the equation), and must be approximated using numerical methods. In this paper, we simulated data using Euler's method, a numerical method for approximating solutions to ODEs with an initial condition (i.e., the starting value of x). In addition, we note that all *solutions* to the system of differential equations in this paper are constrained to be positive, on the interval [0, ∞). This is because the components in our theory cannot take on negative values (e.g., a person can experience zero suicidal thoughts or very intense suicidal thoughts, but not *negative* suicidal thoughts). To formalize this assumption, we use the following mathematical notation:

$$\frac{dx}{dt} = 2x^+ + 4y^+ \tag{1.6}$$

where any component C with the plus superscript (C^+) denotes $max(0, C)$. While we use the components (with the superscript) to parameterize dynamics on the unconstrained space, we are interested in the dynamics of the components with the superscript (constrained to be positive), and it is these superscripted components (e.g., x^+, y^+) that we simulate and plot in all figures.

Step One: Modeling Stressors

The first component in our model included stressors experienced throughout the course of daily life. A large body of research suggests that fluctuations in stressors are characterized by both predictable trends (e.g., chronic, systemic, or structural stressors) and unpredictable events (e.g., acute, episodic, or transient stressors) (Alvarez et al., 2022; Cohodes et al., 2023; Hammen et al., 2009; Harrell, 2000; McGonagle & Kessler, 1990). To embody this theoretical position, we formalized stressors as geometric Brownian motion, a continuous-time process that includes a deterministic and a stochastic component. As shown below, this equation includes two parameters, where the drift parameter μ governs deterministic trends (i.e., higher values of μ lead to increases in stressors over time) and the volatility parameter σ governs stochastic variability (i.e., higher values of σ result in larger fluctuations in stressors). Thus, this mathematical model formalizes both chronic and acute fluctuations in stressors. In addition, geometric Brownian motion only assumes positive values, reflecting the nature of stressors in the real-world (i.e., they vary over time, but cannot go below zero). In mathematical form, we modeled stressors (S) as:

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t} \tag{2}$$

where W_t represents the Wiener process (Brownian Motion), μ = drift, and σ = volatility.

Figure 2 shows six simulations of stressors over two weeks. Of note, the stochastic component of geometric Brownian motion means that each simulation will generate different dynamics over time, even with the same parameter values. Thus, we simulated six *realizations* of stressors (with identical parameter values) to observe a potential distribution of stressors; these can be interpreted as representing stressors over time for six different people. Together, these realizations produce several stressors-related phenomena, including that stressors have between-person variance and within-person fluctuations, and that fluctuations tend to be relatively low (e.g., representing typical day-to-day stressors), but may also display larger perturbations (e.g., representing traumatic or particularly stressful life events). This mathematical model thus allows us to represent a range of fluctuations in stressors observed in the real world.

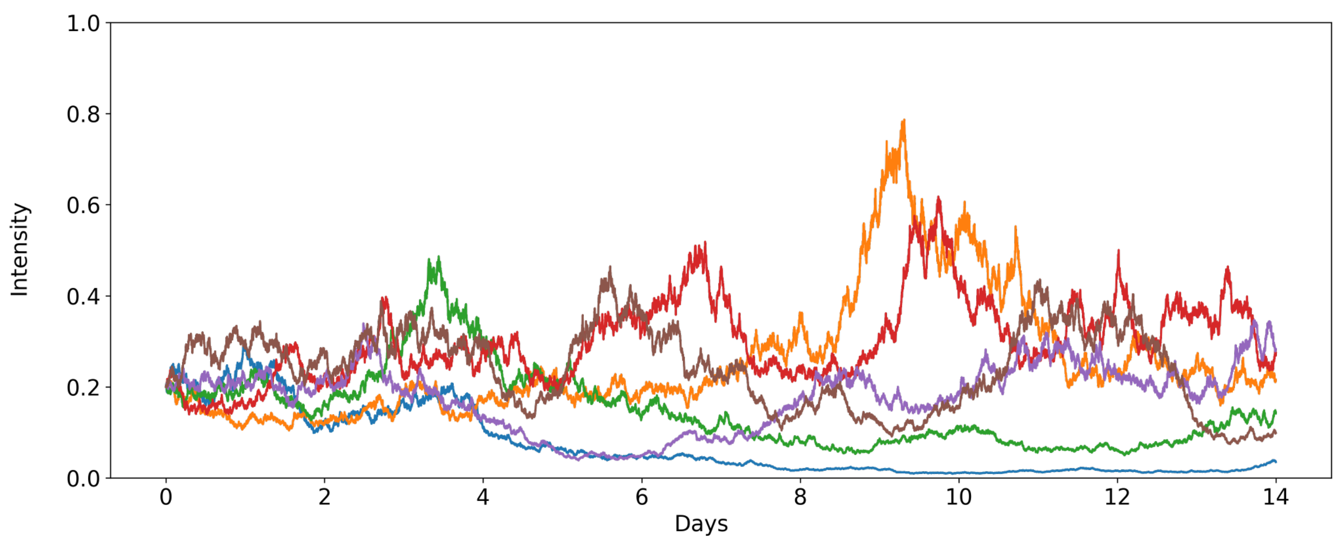

Figure 2. Six realizations of stressors formalized as geometric Brownian motion.

Step Two: Adding Aversive Internal States to the Formal Model

In the General Escape Theory of Suicide, aversive internal states are intentionally represented at a high level to include a broad range of cognitive and affective processes that are experienced as aversive, unpleasant, or negative (e.g., psychological pain, negative affect, hopelessness) (Millner et al., in prep). Aversive internal states are characterized by *within-person fluctuations* in response to external stressors in one's environment (e.g., increased feelings of anger and sadness after social conflict), as well as *between-person differences* in the general tendency to experience aversive internal states (e.g., negative affectivity) (Chaudhury et al., 2017; Vidal Bustamante et al., 2020; Watson & Clark, 1984).

Mathematically, we formalized these within-person fluctuations by modeling aversive internal states as an ODE with a positive effect of stressors (representing the position that increases in external stressors cause increases in aversive internal states over time). To formalize between-person differences in the tendency to experience aversive internal states, we also included a parameter representing a person's *carrying capacity* for aversive internal states, drawing inspiration from ecological population models. In population models, carrying capacity refers to the number of a species (e.g., rabbits) that an environment can support given all available resources (e.g., water, food, shelter). Without predators (e.g., foxes), the population of a species will naturally grow towards its carrying capacity over time. In psychological terms, the "carrying capacity" of aversive internal states can be thought to represent the internal dynamics of aversive internal states that an individual may experience over time, even without considering the effect of other variables (e.g., stressors, escape behaviors). In this way, carrying capacity also allows us to represent individual differences in people's propensity towards aversive internal states, such that, absent other influences, once aversive internal states are present in the system, they will trend towards the individual's *carrying capacity* over time. Thus, in mathematical form, we modeled aversive internal states (A^{+}) as:

$$\frac{dA}{dt} = b_2 A^{+}(K_2 - A^{+}) + a_2 S^{+} \quad (3)$$

where b_2 = the self-feedback loop of aversive internal states, K_2 = carrying capacity of aversive internal states, and a_2 = effect of stressors on change in aversive internal states over time.

To illustrate how flucuations in aversive internal states depend on both their internal dynamics *and* their response to external stressors, we simulated two sets of realizations over two weeks of time, defined by Equations 2 and 3 (see **Figure 3**). As above, within each set of realizations we simulated ata using identical parameter values, such that differences across realizations in each set arise from the stochastic nature of stressors.

In the first set of realizations, we set both the carrying capacity (K_2) and the effect of stressors on aversive internal states (a_2) to be relatively low. As shown in **Figure 3a**, these realizations represent the dynamics of individuals who tend to experience low levels of aversive internal states, even in the face of external stressors (e.g., individuals with low negative affectivity and low emotional reactivity). In the second set of realizations, we set the carrying capacity and the effect of stressors on aversive internal states to be relatively high. As shown in **Figure 3b**, these realizations represent the dynamics of individuals who tend to experience greater levels of aversive internal states, and also react more strongly to external stressors (e.g., individuals with high negative affectivity and high emotional reactivity). Together, these realizations show that fluctuations in aversive internal states are characterized by both within- and between-person fluctuations, depending on one's tendency towards aversive internal states and stressors experienced throughout the course of day-to-day life, thereby providing a plausible minimal model of real-world fluctuations in stressors and aversive internal states over time.

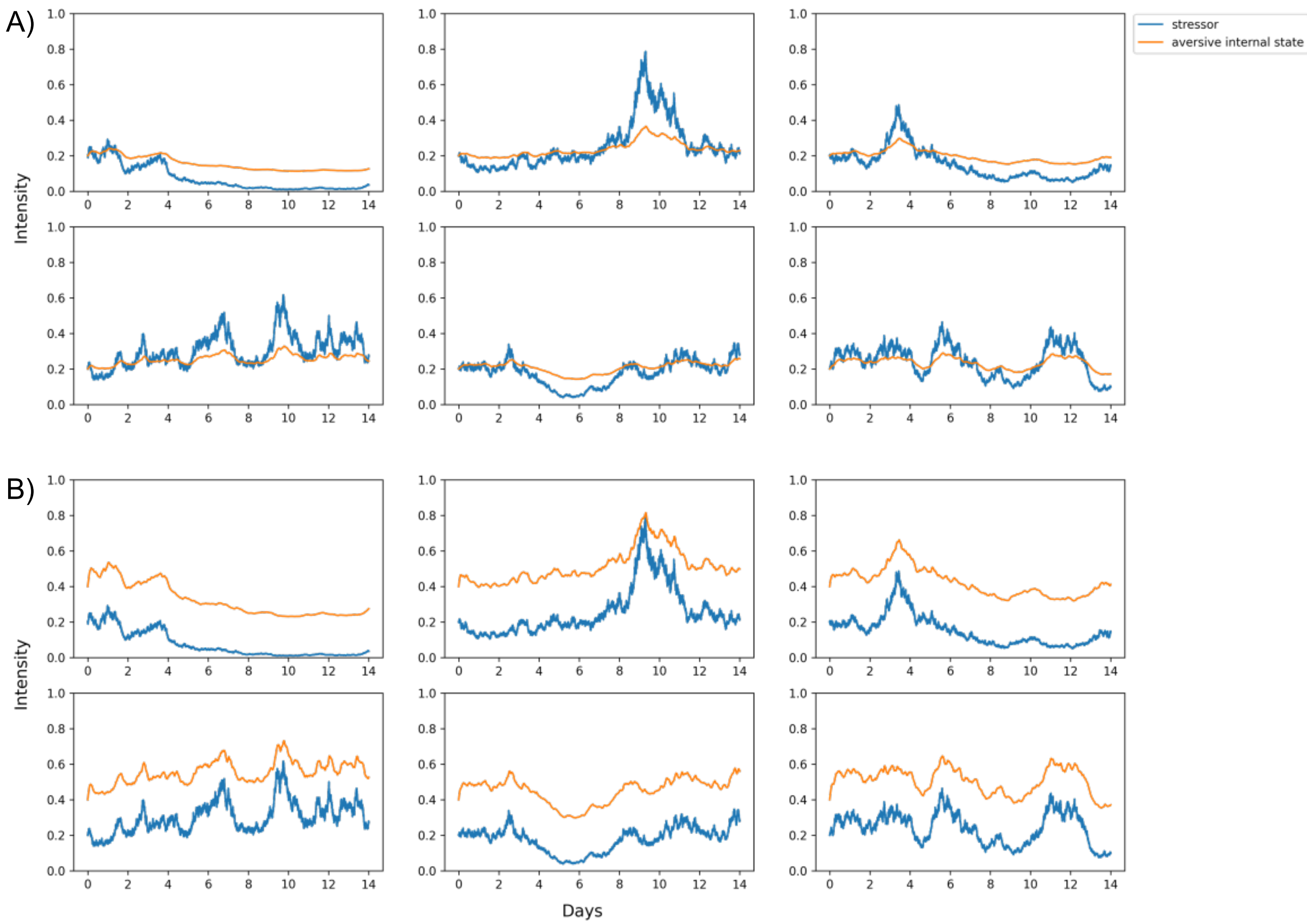

Figure 3. Two sets of realizations of a system including stressors and aversive internal states over two weeks. Panel A) shows fluctuations of a dynamical system with low carrying capacity of aversive internal states, and a weak effect of stressors on change in aversive internal states over time. Panel B) shows fluctuations of a system with higher carrying capacity and a stronger effect of stressors on change in aversive internal states over time. Both panels are simulated with identical fluctuations in stressors.

Step Three: Adding Urge to Escape to the Formal Model

We next added urge to escape into the existing system of equations, along with stressors and aversive internal states. The General Escape Theory of Suicide, as well as other theoretical models of suicide and related problems (e.g., nonsuicidal self-injury, eating disorders) (Juarascio et al., 2017; Linehan, 1987; Lloyd, 2003) and basic research on human responses to aversive contexts (Hartley et al., 2014; Johansen et al., 2011; Öhman & Mineka, 2001), proposes that aversive internal states should increase the urge to escape, but that the urge to escape should also naturally decay over time, as aversive internal states subside. For instance, in Dialectical Behavioral Therapy, clients are taught to "surf the urge" by imagining the urge to escape as waves in the ocean, which can arise quickly and grow in intensity (e.g., due to the effects of stressors and aversive internal states), and can then also subside over time, even without engaging in escape behaviors (Linehan, 1987).

To capture this dynamic of urges to escape that can grow in intensity from aversive internal states while decaying on their own, we modeled the urge to escape as a linear ODE including a negative self-feedback loop (representing natural decay in the urge to escape over time), as well as a positive effect of aversive internal states. Thus, in mathematical form, we modeled the urge to escape (U) as:

$$\frac{dU}{dt} = -c_3U^+ + b_3A^+ \tag{4}$$

where c_3 = self-feedback loop of urge to escape and b_3 = effect of aversive internal states on change in urge to escape over time.

Six realizations of this system (simulated over two weeks of time) defined by Equations 2, 3, and 4 are shown in **Figure 4**. Similar to the simulations above, we observed that fluctuations in the urge to escape over time closely mirror those for aversive internal states and stressors, such that increases in stressors caused momentary increases in both aversive internal states and in the urge to escape. As before, we held parameter values constant across each of the six realizations, such that observed differences across simulations may be attributed to the stochastic nature of stressors, allowing us to visualize a distribution of potential dynamics of this system over time.

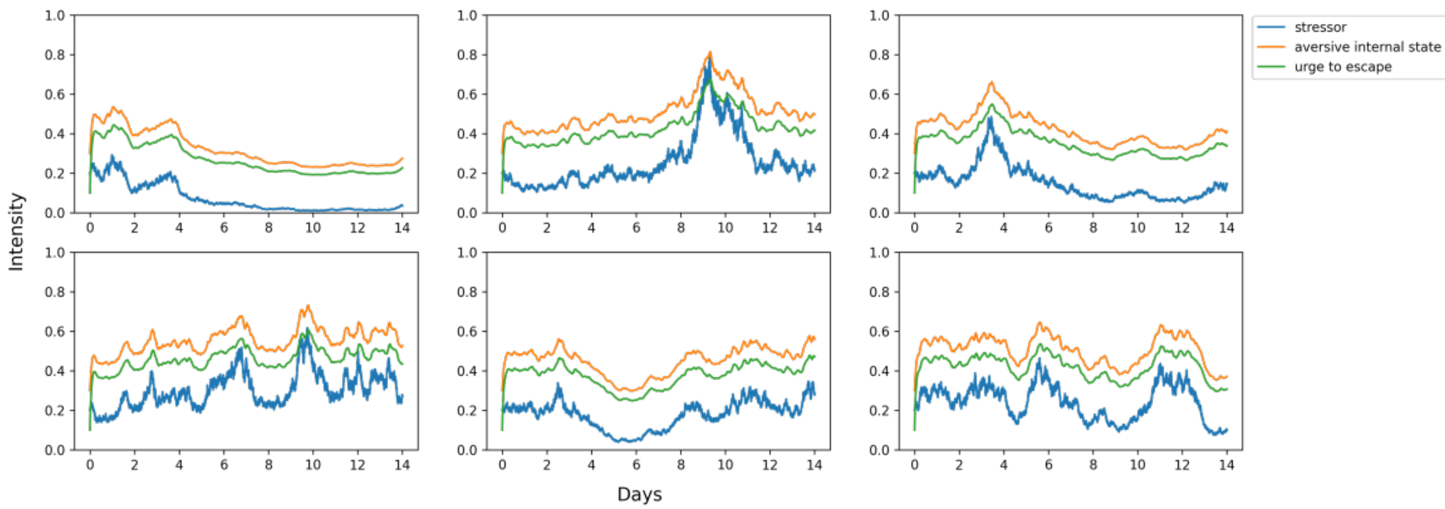

Figure 4. Six realizations of a system including stressors, aversive internal states, and urge to escape over two weeks.

Step Four: Adding Suicidal Thoughts to the Formal Model

Although it is plausible that the urge to escape arises frequently and in approximately linear fashion to aversive internal states, prior research and theory suggest that suicidal thoughts do not emerge in this manner. In other words, it is not the case that small increases in the urge to escape consistently elicit small increases in suicidal thoughts. Rather, ecological momentary assessment (EMA) studies indicate suicidal thoughts are characterized by high zero-inflation. Even in clinical high-risk populations, the most common response to EMA prompts assessing momentary suicidal thoughts is '0' (indicating no suicidal thoughts), even when people report experiencing aversive internal states and an urge to escape. Thus, it must be the case that low (or even high) levels of aversive internal states and escape urges can be present *without* eliciting suicidal thoughts. Further, when suicidal thoughts *do* emerge, they tend to display relatively quick onset and rapid fluctuations over time (Kleiman et al., 2017; Nock et al., 2009; Wang et al., 2021). Therefore, we posited that the effect of urge to escape on suicidal thoughts (T) is non-linear, and implemented this theoretical position using a nonlinear sigmoidal (s-shaped) function:

$$\frac{dT}{dt} = -d_4T^+ + \frac{1}{1 + e^{-c_{41}(U^+ - c_{42})}} \tag{5.1}$$

where d_4 = the self-feedback loop of suicidal thoughts, c_{41}= steepness of the sigmoidal curve, and c_{42} = midpoint of the sigmoidal curve.

Notably, a central theoretical position of the General Escape Theory of Suicide is that suicidal thoughts serve a function: they reduce aversive internal states over time (Kleiman et al., 2018). We implemented this position by incorporating suicidal thoughts as a component (with a negative effect) in the equation governing change in aversive internal states over time. Thus, the revised differential equation for aversive internal states is now:

$$\frac{dA}{dt} = b_2A^+(K_2 - A^+) + a_2S^+ - d_2T^+ \tag{5.2}$$

where d_2 = the effect of suicidal thoughts on change in aversive internal states over time, thereby incorporating the regulating effect of suicidal thoughts.

Six realizations of this dynamical system are presented in **Figure 5.** In these simulations, we observe that no suicidal thoughts emerge in the system when its causal factors (i.e., stressors, aversive internal states, urge to escape) remained relatively low/moderate and stable over time. That is, the system was able to exhibit low and moderate level stressors – and the aversive internal states and urge to escape that accompany those stressors – without exhibiting suicidal thoughts. However, higher fluctuations in these processes resulted in emergence of suicidal thoughts. Of note, even in realizations where suicidal thoughts did emerge, suicidal thoughts remained zero-inflated in the system overall. Thus, the current formal mathematical model was able to successfully reproduce several well-established features concerning the phenomenology of suicidal thoughts, including their relatively quick onset, short duration, and high zero-inflation in time series data (Kleiman et al., 2017; Nock et al., 2009; Wang et al., 2021).

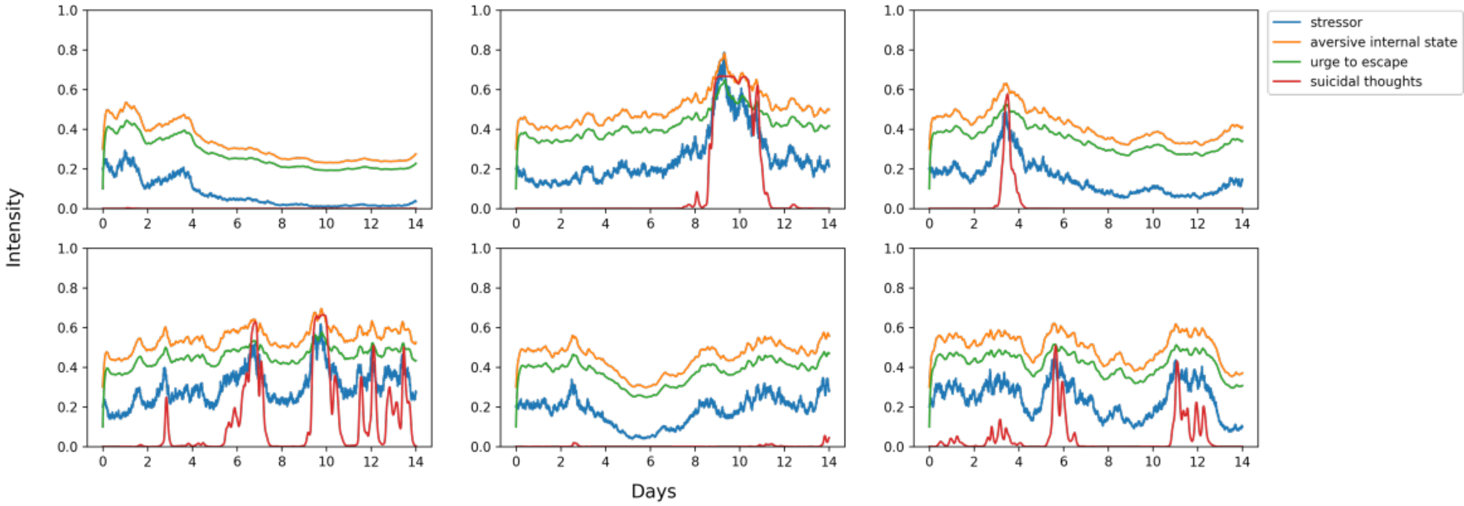

Figure 5. Six realizations of a system including stressors, aversive internal states, urge to escape, and suicidal thoughts over two weeks.

Step Five: Adding Escape Behaviors to the Formal Model

We next added other escape behaviors oriented towards escape of aversive internal states (e.g., alcohol use, nonsuicidal self-injury) into the model. We conceptualized these behaviors as operating similarly to suicidal thoughts, such that they emerge non-linearly with fluctuations in the urge to escape and serve an emotion regulation function by reducing aversive internal states (Adrian et al., 2011; Haynos et al., 2018; Wang et al., 2018). Therefore, similar to suicidal thoughts, we used a sigmoidal function to model escape behaviors (X).

$$\frac{dX}{dt} = -e_5X^+ + \frac{1}{1 + e^{-c_{51}(U^+ - c_{52})}} \tag{6.1}$$

As before, the addition of escape behaviors requires adding its regulating effect into the equation for change in aversive internal states over time. In mathematical form, aversive internal states are now formalized as:

$$\frac{dA}{dt} = b_2A^+(K_2 - A^+) + a_2S^+ - d_2T^+ - e_2X^+ \tag{6.2}$$

where e_2 = the effect of escape behaviors on change in aversive internal states over time, thereby incorporating the regulating effect of escape behaviors.

A central tenet of the General Escape Theory of Suicide (Millner et al., in prep) is that when escape behaviors are successful in reducing aversive internal states, suicidal thoughts should not emerge. Rather, suicidal thoughts should emerge only when an individual experiences persistent aversive internal states, and escape behaviors are no longer effective in regulating them (such that individuals may begin to think about suicide as the "ultimate escape" from their chronic and unrelenting distress). To test if the model indeed produces this phenomena from the General Escape Theory, we ran two sets of simulations for the current system of equations. In both sets of realizations, we held all parameter values constant (as in our previous simulations), except for the e_2 parameter, which represents the ability of escape behaviors to effectively regulate aversive internal states over time. In the first set of simulations, shown in **Figure 6**, we set the e_2 parameter to be relatively high, representing a situation in which escape behaviors successfully regulate aversive internal states. In these simulations, even when stressors became relatively elevated (e.g., in the second and fourth realizations), escape behaviors were sufficiently effective in regulating aversive internal states (and, in turn, the urge to escape), which prevented the emergence of suicidal thoughts over time.

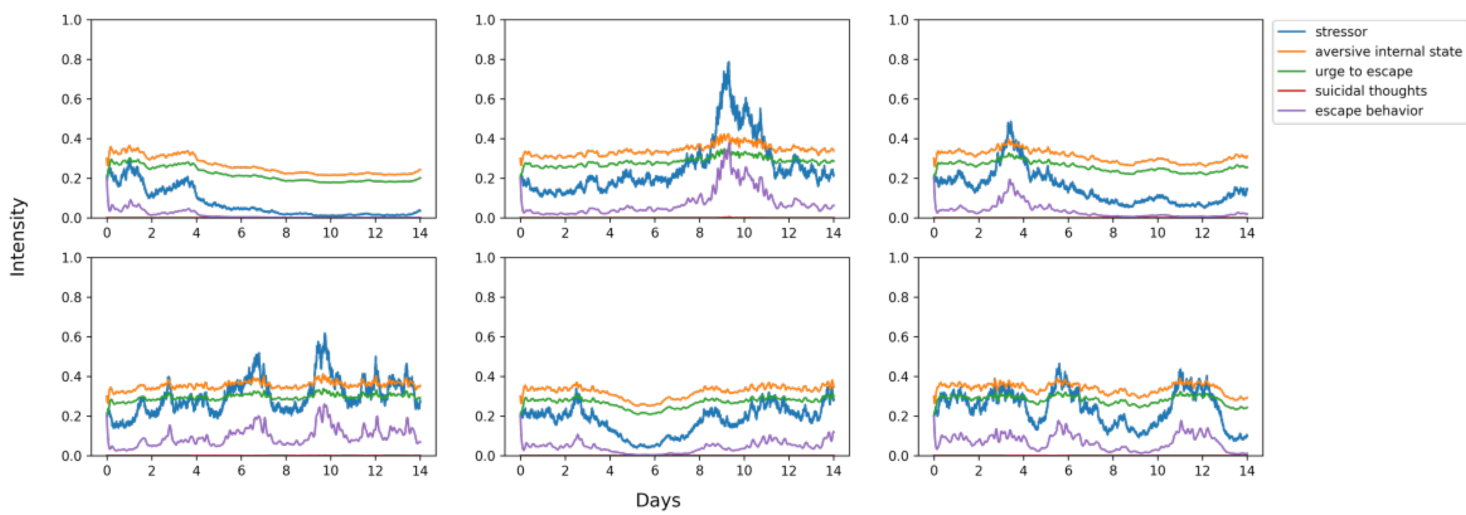

Figure 6. Six realizations of a system in which escape behaviors successfully regulate aversive internal states, preventing the emergence of suicidal thoughts.

In the second set of simulations, we held all other parameter values constant but set e_2 to be relatively low, representing a situation in which *escape behaviors no longer effectively regulated aversive internal states*. As a result, in these simulations (see Figure 7), we observed that suicidal thoughts *did* emerge in conditions of elevated stress, because other escape behaviors were no longer effectively reducing aversive internal states. Thus, the present formal model indeed produced this key phenomenon as predicted by the General Escape Theory of Suicide.

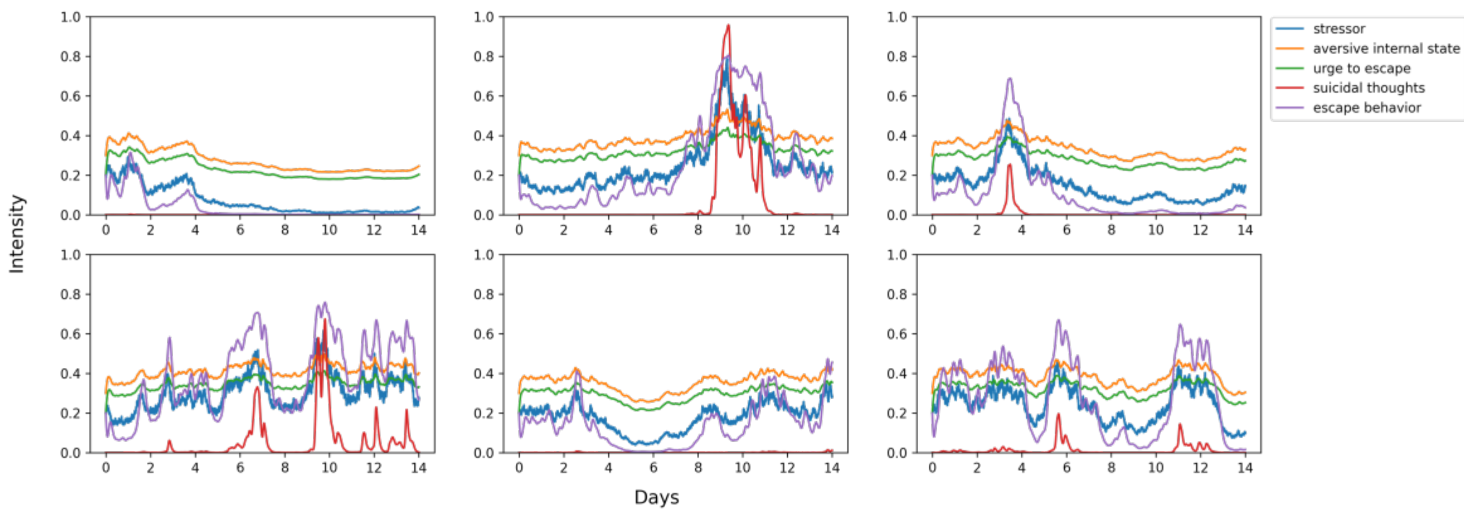

Figure 7. Six realizations of a system in which escape behaviors do *not* regulate negative affect, resulting in the emergence of suicidal thoughts over time.

Step Six: Adding Internal- and External-Focused Strategies to the Formal Model

The final core components of our theory include internal-focused (I) and external-focused (E) strategies, which are theorized to effectively regulate aversive internal states over time. These strategies require individuals to tolerate aversive internal states and enact strategies to reduce them, rather than seeking immediate escape (e.g., via escape behaviors), and intervene on different components of the system: external-focused strategies intervene upon stressors (e.g., meeting with a financial advisor after being fired from a job) and internal-focused strategies intervene on aversive internal states (e.g., engaging in cognitive reappraisal after a fight with a friend) (see **Figure 1**). As with aversive internal states, we conceptualized both internal- and external-focused strategies as including a *carrying capacity* indexing a person's propensity towards engaging in these strategies. Following the General Escape Theory of Suicide, we also included a positive effect of aversive internal states on internal- and external-focused strategies (representing the position that individuals should be motivated to engage in strategies when experiencing aversive internal states), and a negative effect of urge to escape (representing the position that the urge to escape diminishes ability to engage in these strategies) (Nock & Mendes, 2008; Russell et al., 2019). Together, these theoretical positions yield the following set of equations for external- and internal-focused strategies:

$$\frac{dE}{dt} = f_6 E^+ (K_6 - E^+) + b_6 A^+ - c_6 U^+ \tag{7.1}$$

$$\frac{dI}{dt} = g_7 I^+ (K_7 - I^+) + b_7 A^+ - c_7 U^+ \tag{7.2}$$

where f_6= self-feedback loop of external-focused strategies, K_6= carrying capacity of external-focused strategies, b_6= effect of aversive internal states on external-focused strategies, c_6= effect of urge to escape on external-focused strategies, g_7=self-feedback loop of internal-focused strategies, K_7= carrying capacity of internal-focused strategies, b_7= effect of aversive internal states on internal-focused strategies, c_7= effect of urge to escape on internal-focused strategies.

We also added a parameter to reflect the regulating effect of external-focused strategies on stressors over time. The updated and final equation for stressors thus reads:

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t - f_1 E^+} \tag{7.3}$$

Finally, to reflect the regulating effect of emotion-focused strategies on aversive internal states over time, we also added a parameter for aversive internal states. The updated and final equation for aversive internal states is thus:

$$\frac{dA}{dt} = b_2A^+(K_2 - A^+) + a_2S^+ - d_2T^+ - e_2X^+ - g_2I^+ \quad (7.4)$$

The addition of external- and internal-focused strategies completes our model (see **Appendix A** for all final equations). Using this final model, we again performed two sets of simulations to test its ability to produce key theorized phenomena. Specifically, the General Escape Theory of Suicide (Millner et al., in prep) proposes that external- and internal-focused strategies can regulate the system to prevent suicidal thoughts (and other escape behaviors), by effectively regulating stressors and aversive internal states over time. To test the model's ability to produce this behavior, we first set the K_6 and K_7 parameters (representing the carrying capacities for external- and internal-focused strategies, respectively) to be relatively low, representing the dynamics of a system where individuals had a low propensity to engage in external- and internal-focused strategies. Six realizations of this system over two weeks are shown in **Figure 8**. In each of these simulations, we observed relatively low engagement in external- and internal-focused strategies, resulting in persistently high aversive internal states and urges to escape, thus ultimately leading to frequent engagement in escape behaviors and the emergence of suicidal thoughts.

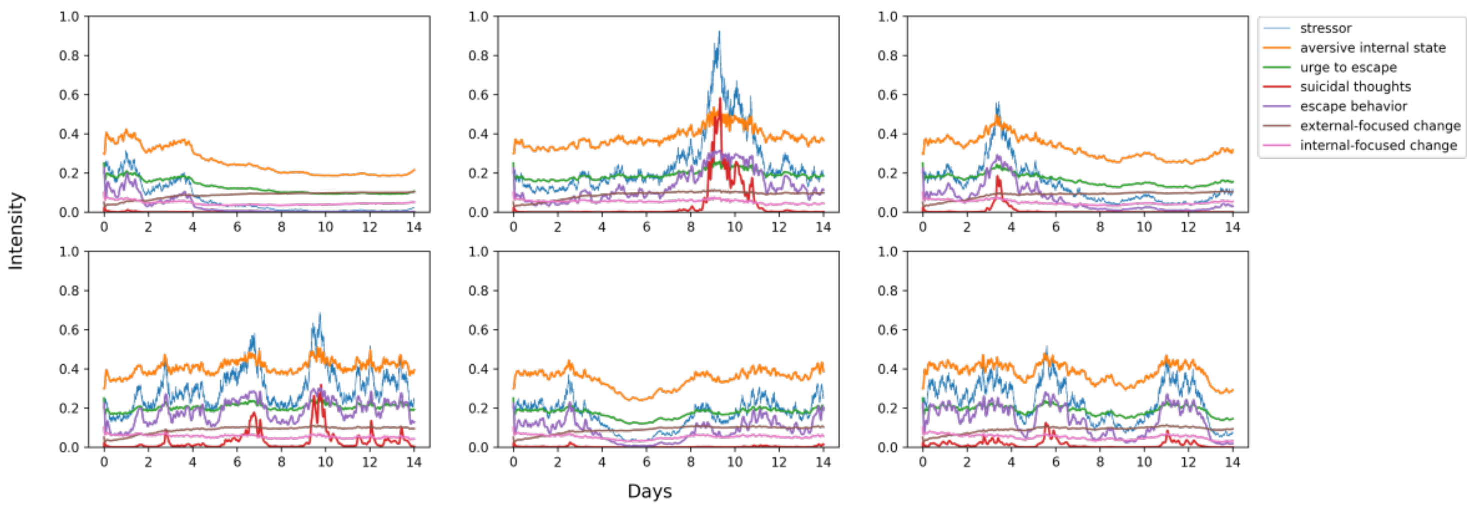

Figure 8. Six realizations with low engagement in external- and internal-focused strategies.

In the second set of simulations, we held all parameter values constant while setting the K_6 and K_6 parameters to be relatively *high* (i.e., increasing the carrying capacities for external- and internal-focused strategies, respectively). We also increased the f_1 and g_2 parameters (i.e., increasing the effectiveness of external-strategies in reducing stressors, and increasing the effectiveness of internal-focused strategies in regulating aversive internal states). These changes in parameters may be thought to represent the effects of an intervention (e.g., helping individuals engage in more frequent and effective coping strategies). Six realizations of this system are illustrated in **Figure 9,** representing the dynamics of a system characterized by high propensity to engage in effective external- and internal-focused strategies. In each of these simulations, we observed higher engagement in external- and internal-focused strategies, which more effectively reduced stressors and aversive internal states over time, leading to low engagement in short-term escape behaviors, and reductions in suicidal thoughts over time. Thus, this formal model was able to produce another key theorized phenomenon of the General Escape Theory of Suicide: that effective use of external- and internal-focused strategies can prevent the emergence of suicidal thoughts over time by reducing the severity of stressors and negative affect.

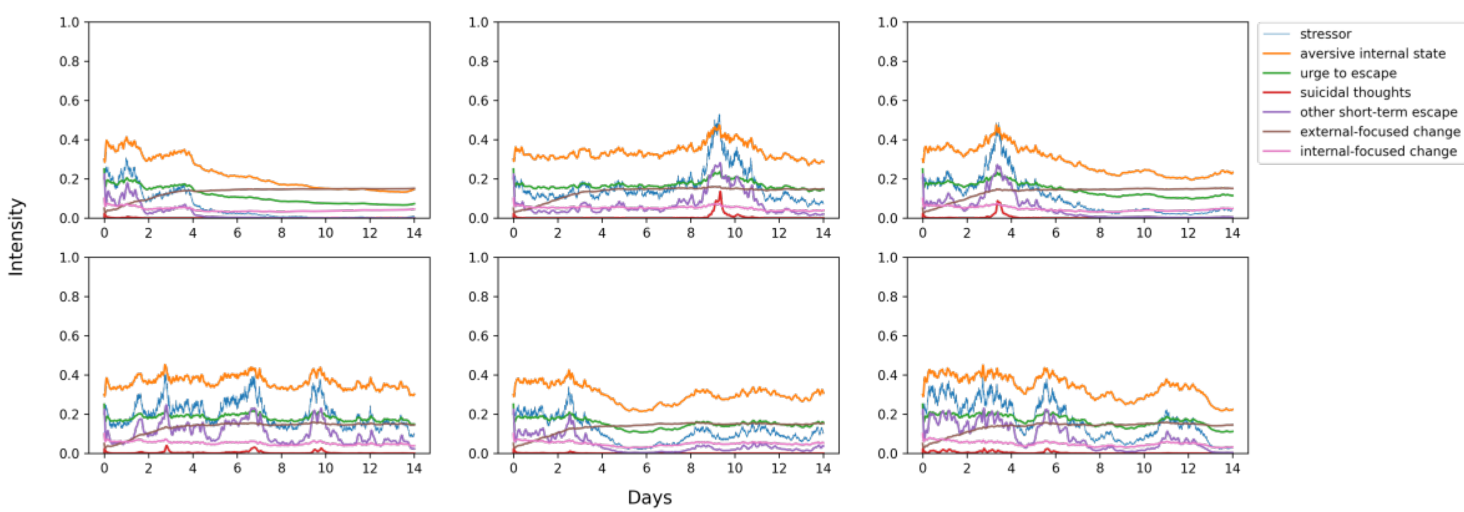

Figure 9. Six realizations of a system with higher and more effective engagement in internal- and external-focused strategies.

Section 3: Studying Suicide as a Complex Dynamical System

In this paper, we presented a formal mathematical model of suicide by formalizing the General Escape Theory of Suicide as a system of stochastic and ordinary differential equations. We believe that the introduction of formal modeling to suicide research represents a meaningful step forward for suicide theory (and theories of psychopathology more broadly), equipping researchers with a set of tools for theory construction provided by mathematical and computational models and – in doing so – illuminating a path by which those theories can be more formally tested and improved. Several noteworthy benefits of theory formalization are illustrated by our modeling efforts here, and warrant further discussion.

Theory Evaluation

Formalizing core components of the General Escape Theory of Suicide as a mathematical model allowed us to *evaluate the explanatory power of our theory by examining its ability to produce robust suicide-related phenomena* (Van Dongen et al., 2022), including that suicidal thoughts are characterized by relatively rapid onset and short duration, high variability and dynamic fluctuations over time, and zero-inflation in time series data (Czyz et al., 2022; Gee et al., 2020; Kivelä et al., 2022; Kleiman et al., 2017; Sedano-Capdevila et al., 2021; Wang et al., 2021). Simulating from our model successfully reproduced these data patterns, allowing us to demonstrate – rather than simply assert – that key features of suicidal thoughts indeed follow from theories rooted in the position that suicide is an escape from aversive internal states. While this does not mean our theory is "correct" (as it is possible that different formalizations of this theory, or different theories entirely, may produce the same data patterns), it does demonstrate that our theory *can* account for these features. This represents a critical improvement over the explanatory power of verbal theories of suicide, which are limited to yielding vague predictions that depend on hidden (implicit) assumptions that often rest in the minds of theorists and can be misunderstood by other researchers seeking to evaluate or use the theory. By formalizing our theory, we have made all assumptions *explicit*, so that our model can be clearly replicated and precisely understood by others who are interested in testing, revising, or building upon it.

Of course, our model is also notably incomplete (as all models are), and continued theory development is needed. For instance, our modeling efforts here were limited by the vast majority of existing suicide research focusing on suicidal *thoughts,* rather than suicidal *behaviors*; this was especially the case for EMA studies, which we relied on to inform our understanding of fluctuations and change in theory components over time (Gee et al., 2020; Kivelä et al., 2022; Sedano-Capdevila et al., 2021). Given the relatively scarce knowledge of how specific psychological processes (and their dynamics) lead to suicide attempts, we focused our initial efforts in this paper on formalizing and modeling the dynamics of a system that gives rise to suicidal thoughts. However, extending this formal model to delineate pathways to suicide attempts is a crucial next step, as we are ultimately interested in understanding and preventing not only suicidal thoughts, but also suicide attempts and death.

Guiding Empirical Research and Discovering Questions

In Section 2, we also used our mathematical and computational model to examine if our theory produced key predictions anticipated by the General Escape Theory of Suicide, including (1) that suicidal thoughts emerge when other behaviors oriented towards short-term escape of aversive internal states are no longer effective at regulating aversive internal states, and (2) that effective engagement in behaviors oriented towards long-term change (i.e., external- and internal-focused strategies) can prevent the emergence of suicidal thoughts (Millner et al., in prep). Simulations from our model indeed produced dynamics in line with the theory's predictions. This is notable, as it lays the foundation for future empirical studies designed to directly test these hypotheses (e.g., that suicidal thoughts emerge with escape behaviors persistently fail to regulate aversive internal states).

Formal models can also help us discover *new* questions to test in empirical studies (Epstein, 2008). For instance, simulations from our model imply that suicidal thoughts always emerge alongside other escape behaviors and that suicidal thoughts cannot be present in the absence of escape behaviors (e.g., see **Figures 7, 8, 9**). In other words, according to this model, *if* someone is experiencing suicidal thoughts, *then* they must also be engaging in other escape behaviors (as well as experiencing aversive internal states, the urge to escape, and stressors). While this may appear to clearly align with clinical intuition, it is a strong prediction and not one that we anticipated *a priori* when developing our verbal theory or formalizing it as a mathematical model. Thus, it raises an interesting empirical question: in the real world, is it always the case that suicidal thoughts occur only in the context of other escape behaviors?

If findings from empirical studies support the robustness of these phenomena (including phenomena that we did and did not anticipate), this would provide strong corroborating evidence in support of our theory. However, if not, then we would have clear evidence for a misalignment between our theory's predictions and dynamics that actually unfold in the real world, providing important information about how the theory might be improved. Importantly, stepping through the process of verbal theory → formal mathematical model → simulations → empirical research, rather than jumping straight from verbal theory → empirical research, allows us to determine if the theory produces what we think it should *before* investing a great deal of resources into empirical studies. This, in turn, strengthens what we can learn from empirical studies and ensures that the resources invested in empirical research have direct implications for theory development.

A “hidden” - and, we believe, crucial - benefit of formalizing theories is that this process can also shed light on potentially overlooked gaps in the literature. For instance, during our formalization process, we found little guidance in the existing literature on how to specify the *form* of relationships between model components (e.g., do aversive internal states cause change in urges to escape over time in a linear or nonlinear manner?). Notably, this imprecision in verbal escape theories only became apparent to us as we attempted to specify each component in our theory as a mathematical model. Given these ambiguities, when specifying model relationships, we followed the modeling principle of “as simple as possible, but no simpler” by relying on linear associations when possible, and extending to nonlinear relationships (e.g., the sigmoidal function employed for modeling suicidal thoughts and escape behaviors) when necessary. Therefore, future work describing the form and *shape* of relationships among theory components, as well as the time scale of their fluctuations, would provide important information to guide continued formal modeling efforts (Millner et al., 2020).

Limitations, Open Questions, and Future Directions

Results from our model should be considered in the context of several limitations and open questions. First, we made the assumption that each simulated time step in our model represented one minute of time (and thus we simulated data over 20,160 time steps, representing two weeks of time). However, it is possible that the simulated dynamics represent fluctuations in stressors, suicidal thoughts, escape behaviors, and other theory components over a different period of time. Indeed, other researchers and clinicians with expertise in suicide may examine our simulated data dynamics from the section above and reasonably disagree on the timescale of observed fluctuations. Critically, we believe that this illustrates a key *benefit* of developing and simulating from such formal theories: by making our theory explicit through mathematical and computational modeling, we can directly observe what we do and do not know, and use this to identify a clear path forward. For instance, future research comparing simulated data to real-time monitoring data collected at different frequencies (e.g., weekly surveys vs. daily diary studies vs. EMA) could provide important information on the timescale of suicidal thinking (and the other constructs in our model) to further evaluate our theory and guide revisions to it.

Second, suicide (and psychopathology more broadly) is inherently complex, and this proposed model is far too simple to accurately represent the system of factors that interact to produce suicidal thoughts and behaviors. For instance, there are many factors that we intentionally did not include in the model, but certainly play an important role in suicide (e.g., positive affect). At the same time, the model we built included a larger number of nodes (7) and parameters (31) than are typically included in formal theoretical models. Most well-known mathematical models in other areas of science typically contain far fewer nodes and parameters; for instance, the Lotka-Volterra model (from ecology) contains only two nodes and four parameters, and the Susceptible-Infected-Recovered (SIR) model (from epidemiology) contains only three nodes and two parameters. The already-large number of parameters in our model makes a comprehensive search of the parameter space computationally infeasible, and thus renders thorough exploration of various combinations of parameter values to be quite challenging. Indeed, as the model grew in complexity, it became increasingly sensitive to even very small changes in parameter values, and we found very little guidance from existing suicide theories or empirical research to guide the selection of parameter values.

Finally, we evaluated the current model via computer simulations, by assessing its ability to produce robust suicide-related phenomena, as well as predictions following from the General Escape Theory of Suicide. A critical next step for this research program involves grounding the model's parameter values in data (i.e., fitting the model to data), and comparing model simulations to empirical data to further evaluate the model's explanatory power. For instance, EMA data would naturally lend themselves to direct comparisons with the time series data we simulated here. Identifying discrepancies between data generated by our theory and data observed in the real world could inform revisions to the formal theory, towards an iterative cycle of theory development, evaluation, and refinement. This is another key benefit of formal vs. verbal theories: formal theories can be modified based on direct comparisons of simulations and empirical data, whereas verbal theories tend to remain stagnant once proposed in the literature (as they are neither strongly supported nor strongly refuted), until they are ultimately replaced by an entirely new verbal theory (Meehl, 1967, 1990b). By making our theory and all its assumptions explicit, open, and reproducible, our hope is that the preliminary model presented here can be continually evaluated and refined by multiple research groups, to build a more robust representation of the complex dynamical system of suicide with each iteration.

Conclusion

Our model represents a first step in the process of building a formal mathematical theory of suicide and offers an exciting new direction in suicide research. Although there is much work to be done, we are enthusiastic about the potential for mathematical and computational modeling methods to generate novel insights on mechanisms and potential treatment targets for suicidal thoughts and behaviors. Specifically, if we have a well-specified model that can reproduce many known and robust real-world phenomena of suicidal thoughts and behaviors, we can then directly intervene on components of the model (e.g., by changing parameter values, adding or removing nodes, etc.) to test the effects of an "intervention," and gain insight on *potential* intervention effects. Given that intervention research is highly cost- and time-intensive, this could be used to explore innovative intervention targets and strategies (and gather pilot data) before designing and conducting a clinical trial with real-world participants.

In addition to generating insights on mechanisms and potential treatment targets, a formal model of suicidal thoughts and behaviors could be a useful clinical tool. For instance, psychoeducation components of many treatment manuals include explaining the short-term effects of engaging in escape behaviors (e.g., “emotion-driven behaviors” in the Unified Protocol) (Barlow et al., 2010). In addition, many modules of various treatment manuals include helping the patient understand the function of a particular problem behavior and identify alternative strategies for responding to negative affect (e.g., behavior chain analysis in Dialectical Behavioral Therapy) (Linehan, 2014). A clinical tool that provides patients with direct “evidence” to support these therapeutic concepts, as well as the ability to explore and observe the effects of different strategies (e.g., short-term escape vs. long-term change behaviors), could help increase acceptability and engagement with the treatment.

We hope that this paper illustrates the promise of formalizing theories as mathematical and computational models, and marks the beginning of a new direction for suicide research: the development of formal theories to guide (and be guided by) rigorous empirical studies, to meaningfully advance our understanding, prediction, and prevention of suicide as a complex dynamical system.

References

Adrian, M., Zeman, J., Erdley, C., Lisa, L., & Sim, L. (2011). Emotional dysregulation and interpersonal difficulties as risk factors for nonsuicidal self-injury in adolescent girls. *Journal of Abnormal Child Psychology*, *39*(3), 389–400.

Alvarez, K., Polanco-Roman, L., Samuel Breslow, A., & Molock, S. (2022). Structural Racism and Suicide Prevention for Ethnoracially Minoritized Youth: A Conceptual Framework and Illustration Across Systems. *American Journal of Psychiatry*, *179*(6), 422–433. https://doi.org/10.1176/appi.ajp.21101001

Barlow, D. H., Farchione, T. J., Fairholme, C. P., Ellard, K. K., Boisseau, C. L., Allen, L. B., & May, J. T. E. (2010). *Unified protocol for transdiagnostic treatment of emotional disorders: Therapist guide*. Oxford University Press.

Baumeister, R. F. (1990). Suicide as escape from self. *Psychological Review*, *97*(1), 90–113. https://doi.org/10.1037/0033-295X.97.1.90

Bentley, K. H., Coppersmith, D. L., Kleiman, E. M., Nook, E. C., Mair, P., Millner, A. J., Reid-Russell, A., Wang, S. B., Fortgang, R. G., Stein, M. B., Beck, S., Huffman, J. C., & Nock, M. K. (2021). Do Patterns and Types of Negative Affect During Hospitalization Predict Short-Term Post-Discharge Suicidal Thoughts and Behaviors? *Affective Science*, *2*(4), 484–494. https://doi.org/10.1007/s42761-021-00058-6

Boccara, N. (2010). *Modeling Complex Systems*. Springer. https://doi.org/10.1007/978-1-4419-6562-2

Borsboom, D., van der Maas, H. L., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory construction methodology: A practical framework for building theories in psychology. *Perspectives on Psychological Science*, *16*(4), 756–766.

Campbell, S. L., & Haberman, R. (2011). *Introduction to differential equations with dynamical systems*. Princeton university press.

CDC. (2021, September 7). *Facts About Suicide*. https://www.cdc.gov/suicide/facts/index.html

Chaudhury, S. R., Galfalvy, H., Biggs, E., Choo, T.-H., Mann, J. J., & Stanley, B. (2017). Affect in response to stressors and coping strategies: An ecological momentary assessment study of borderline personality disorder. *Borderline Personality Disorder and Emotion Dysregulation*, *4*(1), 8. https://doi.org/10.1186/s40479-017-0059-3

Cohodes, E. M., McCauley, S., Pierre, J. C., Hodges, H. R., Haberman, J. T., Santiuste, I., Rogers, M. K., Wang, J., Mandell, J. D., & Gee, D. G. (2023). Development and validation of the Dimensional Inventory of Stress and Trauma Across the Lifespan (DISTAL): A novel assessment tool to facilitate the dimensional study of psychobiological sequelae of exposure to adversity. *Developmental Psychobiology*, *65*(4), e22372. https://doi.org/10.1002/dev.22372

Czyz, E. K., Koo, H. J., Al-Dajani, N., Kentopp, S. D., Jiang, A., & King, C. A. (2022). Temporal profiles of suicidal thoughts in daily life: Results from two mobile-based monitoring studies with high-risk adolescents. *Journal of Psychiatric Research*, *153*, 56–63. https://doi.org/10.1016/j.jpsychires.2022.06.050

Epstein, J. M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, *11*(4), 12.

Farlow, S. J. (2006). *An Introduction to Differential Equations and Their Applications*. Courier Corporation.

Fisher, D. N., & Pruitt, J. N. (2020). Insights from the study of complex systems for the ecology and evolution of animal populations. *Current Zoology*, *66*(1), 1–14. https://doi.org/10.1093/cz/zoz016

Fortgang, R. G., & Nock, M. K. (2021). Ringing the Alarm on Suicide Prevention: A Call to Action. *Psychiatry*, *84*(2), 192–195. https://doi.org/10.1080/00332747.2021.1907871

Fox, K. R., Huang, X., Guzmán, E. M., Funsch, K. M., Cha, C. B., Ribeiro, J. D., & Franklin, J. C. (2020). Interventions for suicide and self-injury: A meta-analysis of randomized controlled trials across nearly 50 years of research. *Psychological Bulletin*. https://doi.org/10.1037/bul0000305

Franklin, J. C., Ribeiro, J. D., Fox, K. R., Bentley, K. H., Kleiman, E. M., Huang, X., Musacchio, K. M., Jaroszewski, A. C., Chang, B. P., & Nock, M. K. (2017). Risk factors for suicidal thoughts and behaviors: A meta-analysis of 50 years of research. *Psychological Bulletin*, *143*(2), 187–232. https://doi.org/10.1037/bul0000084

Fried, E. I. (2020). Lack of Theory Building and Testing Impedes Progress in The Factor and Network Literature. *Psychological Inquiry*, *31*(4), 271–288. https://doi.org/10.1080/1047840X.2020.1853461

Gee, B. L., Han, J., Benassi, H., & Batterham, P. J. (2020). Suicidal thoughts, suicidal behaviours and self-harm in daily life: A systematic review of ecological momentary assessment studies. *DIGITAL HEALTH*, *6*, 2055207620963958. https://doi.org/10.1177/2055207620963958

Gee, D. G., & Casey, B. J. (2015). The impact of developmental timing for stress and recovery. *Neurobiology of Stress*, *1*, 184–194. https://doi.org/10.1016/j.ynstr.2015.02.001

Guest, O., & Martin, A. E. (2020). *How computational modeling can force theory building in psychological science*. PsyArXiv. https://doi.org/10.31234/osf.io/rybh9

Gunn III, J. F. (2014). Suicide as escape: Baechler, Shneidman, and Baumeister. *Theories of Suicide: Past, Present, and Future*, 9–22.

Hammen, C., Kim, E. Y., Eberhart, N. K., & Brennan, P. A. (2009). Chronic and acute stress and the prediction of major depression in women. *Depression and Anxiety*, *26*(8), 718–723. https://doi.org/10.1002/da.20571

Harrell, S. P. (2000). A Multidimensional Conceptualization of Racism-Related Stress: Implications for the Well-Being of People of Color. *American Journal of Orthopsychiatry*, *70*(1), 42–57. https://doi.org/10.1037/h0087722

Hartley, C. A., Gorun, A., Reddan, M. C., Ramirez, F., & Phelps, E. A. (2014). Stressor controllability modulates fear extinction in humans. *Neurobiology of Learning and Memory*, *113*, 149–156. https://doi.org/10.1016/j.nlm.2013.12.003

Haynos, A. F., Wang, S. B., & Fruzzetti, A. E. (2018). Restrictive eating is associated with emotion regulation difficulties in a non-clinical sample. *Eating Disorders*, *26*(1), 5–12. https://doi.org/10.1080/10640266.2018.1418264

Hennings, J. M. (2020). Function and Psychotherapy of Chronic Suicidality in Borderline Personality Disorder: Using the Reinforcement Model of Suicidality. *Frontiers in Psychiatry*, *11*. https://www.frontiersin.org/articles/10.3389/fpsyt.2020.00199

Johansen, J. P., Cain, C. K., Ostroff, L. E., & LeDoux, J. E. (2011). Molecular Mechanisms of Fear Learning and Memory. *Cell*, *147*(3), 509–524. https://doi.org/10.1016/j.cell.2011.10.009

Joiner, T. E. (2005). *Why people die by suicide*. Harvard University Press.

Juarascio, A. S., Manasse, S. M., Schumacher, L., Espel, H., & Forman, E. M. (2017). Developing an acceptance-based behavioral treatment for binge eating disorder: Rationale and challenges. *Cognitive and Behavioral Practice*, *24*(1), 1–13.

Kazdin, A. E. (2012). *Behavior Modification in Applied Settings: Seventh Edition*. Waveland Press.

Kivelä, L., van der Does, W. A. J., Riese, H., & Antypa, N. (2022). Don't Miss the Moment: A Systematic Review of Ecological Momentary Assessment in Suicide Research. *Frontiers in Digital Health*, *4*. https://www.frontiersin.org/articles/10.3389/fdgth.2022.876595

Kleiman, E. M., Coppersmith, D. D. L., Millner, A. J., Franz, P. J., Fox, K. R., & Nock, M. K. (2018). Are suicidal thoughts reinforcing? A preliminary real-time monitoring study on the potential affect regulation function of suicidal thinking. *Journal of Affective Disorders*, *232*, 122–126. https://doi.org/10.1016/j.jad.2018.02.033

Kleiman, E. M., Turner, B. J., Fedor, S., Beale, E. E., Huffman, J. C., & Nock, M. K. (2017). Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies. *Journal of Abnormal Psychology*, *126*(6), 726–738. https://doi.org/10.1037/abn0000273

Klonsky, E. D., & May, A. M. (2015). The three-step theory (3ST): A new theory of suicide rooted in the "ideation-to-action" framework. *International Journal of Cognitive Therapy*, *8*(2), 114–129.

Kuehn, K. S., Dora, J., Harned, M. S., Foster, K. T., Song, F., Smith, M. R., & King, K. M. (2022). A meta-analysis on the affect regulation function of real-time self-injurious thoughts and behaviours. *Nature Human Behaviour*, *6*(7), Article 7. https://doi.org/10.1038/s41562-022-01340-8

Lass, A. N. S., Winer, E. S., Collins, A. C., & Rokke, P. D. (2020). The Associations among Distress Tolerance, Unhelpful Coping Behaviors, and Symptoms of Depression: A Network Analysis. *Cognitive Therapy and Research*, *44*(6), 1077–1090. https://doi.org/10.1007/s10608-020-10131-2

Linehan, M. (2014). *DBT? Skills training manual*. Guilford Publications.

Linehan, M. M. (1987). Dialectical behavior therapy for borderline personality disorder. Theory and method. *Bulletin of the Menninger Clinic*, *51*(3), 261–276.

Lloyd, A. (2003). Urge surfing. *Cognitive Behavior Therapy: Applying Empirically Supported Techniques in Your Practice*, 451–455.

McGonagle, K. A., & Kessler, R. C. (1990). Chronic stress, acute stress, and depressive symptoms. *American Journal of Community Psychology*, *18*(5), 681–706. https://doi.org/10.1007/BF00931237

McLaughlin, K. A., & Hatzenbuehler, M. L. (2009). Mechanisms Linking Stressful Life Events and Mental Health Problems in a Prospective, Community-Based Sample of Adolescents. *Journal of Adolescent Health*, *44*(2), 153–160. https://doi.org/10.1016/j.jadohealth.2008.06.019

Meehl, P. E. (1967). Theory-testing in psychology and physics: A methodological paradox. Philosophy of Science. *Meehl, Research*, 103–115.

Meehl, P. E. (1990a). Appraising and Amending Theories: The Strategy of Lakatosian Defense and Two Principles that Warrant It. *Psychological Inquiry*, *1*(2), 108–141. https://doi.org/10.1207/s15327965pli0102_1

Meehl, P. E. (1990b). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, *66*(1), 195–244.

Millner, A. J., Robinaugh, D. J., & Nock, M. K. (2020). Advancing the Understanding of Suicide: The Need for Formal Theory and Rigorous Descriptive Research. *Trends in Cognitive Sciences*, *24*(9), 704–716. https://doi.org/10.1016/j.tics.2020.06.007

Mongin, D., Uribe, A., Cullati, S., & Courvoisier, D. S. (20220801). A tutorial on ordinary differential equations in behavioral science: What does physics teach us? *Psychological Methods*. https://doi.org/10.1037/met0000517

Muthukrishna, M., & Henrich, J. (2019). A problem in theory. *Nature Human Behaviour*, *3*(3), 221–229. https://doi.org/10.1038/s41562-018-0522-1

Nock, M. K., & Mendes, W. B. (2008). Physiological arousal, distress tolerance, and social problem-solving deficits among adolescent self-injurers. *Journal of Consulting and Clinical Psychology*, *76*(1), 28–38. https://doi.org/10.1037/0022-006X.76.1.28

Nock, M. K., & Prinstein, M. J. (2004). A Functional Approach to the Assessment of Self-Mutilative Behavior. *Journal of Consulting and Clinical Psychology*, *72*(5), 885–890. https://doi.org/10.1037/0022-006X.72.5.885

Nock, M. K., Prinstein, M. J., & Sterba, S. K. (2009). Revealing the form and function of self-injurious thoughts and behaviors: A real-time ecological assessment study among adolescents and young adults. *Journal of Abnormal Psychology*, *118*(4), 816–827. https://doi.org/10.1037/a0016948

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, *108*(3), 483–522. https://doi.org/10.1037/0033-295X.108.3.483

Olthof, M., Hasselman, F., Oude Maatman, F., Bosman, A. M., & Lichtwarck-Aschoff, A. (2023). Complexity theory of psychopathology. *Journal of Psychopathology and Clinical Science*, *132*(3), 314.

Ottino, J. M. (2003). Complex systems. *American Institute of Chemical Engineers. AIChE Journal*, *49*(2), 292.

Peterson, C. M., Davis-Becker, K., & Fischer, S. (2014). Interactive role of depression, distress tolerance and negative urgency on non-suicidal self-injury. *Personality and Mental Health*, *8*(2), 151–160. https://doi.org/10.1002/pmh.1256

Ricardo, H. J. (2020). *A modern introduction to differential equations*. Academic Press.

Robinaugh, D. J., Haslbeck, J. M., Ryan, O., Fried, E. I., & Waldorp, L. J. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction. *Perspectives on Psychological Science*, *16*(4), 725–743.

Russell, B. S., Lincoln, C. R., & Starkweather, A. R. (2019). Distress Tolerance Intervention for Improving Self-Management of Chronic Conditions: A Systematic Review. *Journal of Holistic Nursing*, *37*(1), 74–86. https://doi.org/10.1177/0898010118777327

Scheffer, M., Bascompte, J., Brock, W. A., Brovkin, V., Carpenter, S. R., Dakos, V., Held, H., van Nes, E. H., Rietkerk, M., & Sugihara, G. (2009). Early-warning signals for critical transitions. *Nature*, *461*(7260), 53–59. https://doi.org/10.1038/nature08227

Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., & Walker, B. (2001). Catastrophic shifts in ecosystems. *Nature*, *413*(6856), 591–596. https://doi.org/10.1038/35098000

Sedano-Capdevila, A., Porras-Segovia, A., Bello, H. J., Baca-García, E., & Barrigon, M. L. (2021). Use of ecological momentary assessment to study suicidal thoughts and behavior: A systematic review. *Current Psychiatry Reports*, *23*(7), 1–17.

Shneidman, E. S. (1993). Commentary: Suicide as psychache. *Journal of Nervous and Mental Disease*, *181*(3), 145–147. https://doi.org/10.1097/00005053-199303000-00001

Siegenfeld, A. F., & Bar-Yam, Y. (2020). An Introduction to Complex Systems Science and Its Applications. *Complexity*, *2020*, 1–16. https://doi.org/10.1155/2020/6105872

Smaldino, P. E. (2017). Models Are Stupid, and We Need More of Them. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational Social Psychology* (1st ed., pp. 311–331). Routledge. https://doi.org/10.4324/9781315173726-14

Van Dongen, N. N. N., Van Bork, R., Finnemann, A., Van Der Maas, H., Robinaugh, D., Haslbeck, J. M. B., De Ron, J., Sprenger, J., & Borsboom, D. (2022). *Productive Explanation: A Framework for Evaluating Explanations in Psychological Science* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/qd69g

Van Nes, E. H., & Scheffer, M. (2007). Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *The American Naturalist*, *169*(6), 738–747.

van Rooij, I., & Baggio, G. (2020). Theory development requires an epistemological sea change. *Psychological Inquiry*, *31*(4), 321–325.

Vidal Bustamante, C. M., Rodman, A. M., Dennison, M. J., Flournoy, J. C., Mair, P., & McLaughlin, K. A. (2020). Within-person fluctuations in stressful life events, sleep, and anxiety and depression symptoms during adolescence: A multiwave prospective study. *Journal of Child Psychology and Psychiatry*, *61*(10), 1116–1125. https://doi.org/10.1111/jcpp.13234

Wang, S. B., & Borders, A. (2018). The unique effects of angry and depressive rumination on eating-disorder psychopathology and the mediating role of impulsivity. *Eating Behaviors*, *29*, 41–47. https://doi.org/10.1016/j.eatbeh.2018.02.004

Wang, S. B., Coppersmith, D. D. L., Kleiman, E. M., Bentley, K. H., Millner, A. J., Fortgang, R., Mair, P., Dempsey, W., Huffman, J. C., & Nock, M. K. (2021). A Pilot Study Using Frequent Inpatient Assessments of Suicidal Thinking to Predict Short-Term Postdischarge Suicidal Behavior. *JAMA Network Open*, *4*(3), e210591. https://doi.org/10.1001/jamanetworkopen.2021.0591

Wang, S. B., Fox, K. R., Boccagno, C., Hooley, J. M., Mair, P., Nock, M. K., & Haynos, A. F. (2021). Functional assessment of restrictive eating: A three-study clinically heterogeneous and transdiagnostic investigation. *Journal of Abnormal Psychology*, *130*(7), 761.

Wang, S. B., Pisetsky, E. M., Skutch, J. M., Fruzzetti, A. E., & Haynos, A. F. (2018). Restrictive eating and nonsuicidal self-injury in a nonclinical sample: Co-occurrence and associations with emotion dysregulation and interpersonal problems. *Comprehensive Psychiatry*, *82*, 128–132. https://doi.org/10.1016/j.comppsych.2018.02.005

Wang, S. B., Van Genugten, R., Yacoby, Y., Pan, W., Bentley, K., Bird, S., Buonopane, R., Christie, A., Daniel, M., DeMarco, D., Haim, A., Follet, L., Fortgang, R., Kelly, F., Kleiman, E., Millner, A., Obi-Obasi, O., Onnela, J.-P., Ramlal, N., … Nock, M. (2023). *Idiographic Prediction of Suicidal Thoughts: Building Personalized Machine Learning Models with Real-Time Monitoring Data* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/qh4jk

Watson, D., & Clark, L. A. (1984). Negative affectivity: The disposition to experience aversive emotional states. *Psychological Bulletin*, *96*(3), 465.

WHO. (2018). *National suicide prevention strategies: Progress, examples and indicators*.

Appendix A. Full list of the system of stochastic and ordinary differential equations formalizing core components of the General Escape Theory of Suicide.

$$S_t = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t - f_1 E^+}$$

$$\frac{dA}{dt} = b_2 A^+ (K_2 - A^+) + a_2 S^+ - d_2 T^+ - e_2 X^+ - g_2 I^+$$

$$\frac{dU}{dt} = -c_3 U^+ + b_3 A^+$$

$$\frac{dT}{dt} = -d_4 T^+ + \frac{1}{1 + e^{-c_{41}(U^+ - c_{42})}}$$

$$\frac{dX}{dt} = -e_5 X^+ + \frac{1}{1 + e^{-c_{51}(U^+ - c_{52})}}$$

$$\frac{dE}{dt} = f_6 E^+ (K_6 - E^+) + b_6 A^+ - c_6 U^+$$

$$\frac{dI}{dt} = g_7 I^+ (K_7 - I^+) + b_7 A^+ - c_7 U^+$$

where S = stressors, A = aversive internal states, U = urge to escape, T = suicidal thoughts, X = escape behaviors, E = external-focused strategies, and I = internal-focused strategies. As described in the manuscript, all components C with the plus superscript denote $max(0, C)$. While we use components (with the superscript) to parameterize dynamics on the unconstrained space, we are interested in the dynamics of the components with the superscript (constrained to be positive), and it is these superscripted components (e.g., A^+, U^+, T^+) that we simulate and plot in all figures.