Dynamic Modeling, System Identification, and Control Engineering Approaches for Designing Optimized and Perpetually Adaptive Behavioral Health Interventions

by

Mohammad T. Freigoun

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2021 by the
Graduate Supervisory Committee:

Gregory B. Raupp, Co-Chair
Konstantinos S. Tsakalis, Co-Chair
Andreas S. Spanias
Erica S. Forzani
Christopher L. Muhich

ARIZONA STATE UNIVERSITY

August 2021

ABSTRACT

Behavior-driven obesity has become one of the most challenging global epidemics since the 1990s, and is presently associated with the leading causes of death in the U.S. and worldwide, including diabetes, cardiovascular disease, strokes, and some forms of cancer. The use of system identification and control engineering principles in the design of novel and perpetually adaptive behavioral health interventions for promoting physical activity and healthy eating has been the central theme in many recent contributions. However, the absence of experimental studies specifically designed with the purpose of developing control-oriented behavioral models has restricted prior efforts in this domain to the use of hypothetical simulations to demonstrate the potential viability of these interventions. In this dissertation, the use of first-of-a-kind, real-life experimental results to develop dynamic, participant-validated behavioral models essential for the design and evaluation of optimized and adaptive behavioral interventions is examined.

Following an intergenerational approach, the first part of this work aims to develop a dynamical systems model of intrauterine fetal growth with the prime goal of predicting infant birth weight, which has been associated with subsequent childhood and adult-onset obesity. The use of longitudinal input-output data from the "Healthy Mom Zone" intervention study has enabled the estimation and validation of this fetoplacental model. The second part establishes a set of data-driven behavioral models founded on Social Cognitive Theory (SCT). The "Just Walk" intervention experiment, developed at Arizona State University using system identification principles, has lent a unique opportunity to estimate and validate both black-box and semiphysical SCT models for predicting physical activity behavior. Further, this dissertation addresses some of the model estimation challenges arising from the limitations of "Just Walk", including the need for developing nontraditional modeling

approaches for short datasets, as well as delivers a new theoretical and algorithmic framework for structured state-space model estimation that can be used in a broader set of application domains. Finally, adaptive closed-loop intervention simulations of participant-validated SCT models from "Just Walk" are presented using a Hybrid Model Predictive Control (HMPC) control law. A simple HMPC controller reconfiguration strategy for designing both single- and multi-phase intervention designs is proposed.

To the spirit of my mother, Samia.

To my father, Tarig.

To my siblings, Hala, Sara, and Omer.

To Rasitti.

To my beloved fiancé, Douana.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my utmost gratitude to my advisors Professor Gregory B. Raupp and Professor Konstantinos S. Tsakalis for all their time and support, and ultimately for being instrumental in bringing the work of this dissertation to light. I would like to specifically thank them for entrusting me to continue with my research with reasonable autonomy, and provide nothing but constructive feedback, always in the most positive spirit. My special thanks goes to Dr. Paulo Lopes dos Santos whose collaboration and revelations were invaluable for my professional growth in general and for the completion of this work in particular.

Further, I would like to share that it is my greatest honor and privilege to have been taught by Dr. Andreas Spanias, Dr. Erica Forzani and Dr. Christopher Muhich, who also served as members in my dissertation committee, and have their blessings to join the community of scholars. To them I say thank you for your support, valuable insights, and mostly for always putting the interests of your students first. My appreciation to Dr. Bin Mu for serving in my committee before leaving ASU. I am also grateful to the Director of the School for Engineering of Matter, Transport & Energy (SEMTE), Dr. Lenore Dai, the graduate program chair for chemical engineering, Dr. Marylaura Lind Thomas, and the graduate advising manager within SEMTE, Ms. Tiffany Wingerson, for their support throughout my journey in the program. A personal and warm thank you to Ms. Christine Quintero for all her trust and help in navigating through the program, particularly in the most uncertain of times. She will always inspire me to treat all the vulnerable with respect and kindness, and fulfil my job responsibilities with endless dedication and professionalism.

My appreciation goes beyond and includes all co-authors of my publications, with a particular acknowledgment of the guidance from Dr. Daniel Rivera, Dr. Eric Hekler, Dr. Jennifer Savage, and Dr. Danielle Downs during my work in the *Just Walk* and

Healthy Mom Zone projects. A special appreciation to Professor Hans D. Mittelmann who delivered my first and most favorite ASU course of all: APM 523 Optimization (Applied Mathematics).

I also wish to express all the best to my former colleagues, Dr. César Martín and Dr. Penghong Guo, for all the great times and fun conversations we had; thank you for enriching my graduate school experience on many levels. To my dearest friend from ASU, Alicia Magann, who remained in touch even after her graduation: your friendship has always been a source of joy and pride; I sincerely wish you all the best in your professional and personal life.

Last but certainly not least, I would like to all my lifelong friends in Arizona and abroad who continue to inspire me every day. To Mr. John Wayne Mercure, thank you for your generosity. To the dear Nedal Abdeen, thank you for all your support, inspiration, and continued friendship.

Research work presented in this proposal was, in part, funded by the NIH grants R56-HL126799 & R01-HL119245, NSF grant IIS-1449751, and Arizona State University. Opinions expressed in this dissertation are the author's own and do not necessarily reflect the views of any mentioned individual or entity.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

1.1 Motivation

System identification and control are among the most mature fields in engineering that offer rigorous mathematical framework for the characterization of complex dynamical behaviors of causal systems, and ultimately utilize predictive models to establish optimized real-time solutions in a wide array of applications. Beyond process design and other traditional industrial settings, system identification and control research have been expanding and now include novel applications in new areas such as behavioral health and medicine. Despite the ever-growing body of scientific literature and subsequent efforts devoted to public awareness, some of the most pressing challenges in behavioral health remain unresolved. In the United States, according to current reports from Centers for Disease Control and Prevention (CDC), among the top ten public health problems are obesity, HIV, drug abuse, and tobacco use, with obesity strongly tied to other problems in this category such as heart disease and stroke. Over the recent decades, following smoking, the obesity pandemic was the second leading cause of preventable deaths estimated at 300,000 mortalities per year in the United States alone (Flegal *et al.*, 2004). Moreover, being associated with type 2 diabetes, hypertension, heart disease, asthma, depression, and even cancer among other conditions, the United States national direct costs of obesity were estimated at $147 billion in 2008 (Finkelstein *et al.*, 2009; Cawley and Meyerhoefer, 2012). Indirect costs including obesity-related absenteeism have ranged between $3.38 and $6.38 billion (Trogdon *et al.*, 2008).

While it is more difficult to prevent some of the risk factors such as age, family history and genetic factors, regulation of one's habits and behavior concerning physical activity, diet, and tobacco use is indeed more approachable, and can remarkably increase life expectancy. For example, substantial evidence shows that physical activity reduces chronic disease risk (Bauer *et al.*, 2014; Owen *et al.*, 2008; Haskell *et al.*, 2009). With national guidelines suggesting 30-60 minutes of moderate physical activity per day, often in the form of walking (U.S. Department of Health and Human Services, 2008); research furthermore demonstrates a 20-30% reduced risk of breast cancer when guidelines are met (Thune *et al.*, 1997). Despite the well-known benefits of physical activity, a large segment of the U.S. population does not meet these guidelines (Troiano *et al.*, 2008).

In addition to sedentary behavior, Gestational Weight Gain (GWG) and high birth weight have also been independently associated with subsequent childhood and adult-onset obesity, cardiovascular disease, and some forms of cancer (Rich-Edwards *et al.*, 1997; Hillier *et al.*, 2007; Spracklen *et al.*, 2014; Savage *et al.*, 2014; Qiao *et al.*, 2015; O'Neill *et al.*, 2015). While it is known that these associations can impact both women and their offspring, adherence of pregnant women in the United States to guidelines from the Institute of Medicine for appropriate GWG is only at 30%, with over 50% exceeding guidelines toward overweight and morbid obesity (Dudenhausen *et al.*, 2015). These low levels of adherence to recommended guidelines indeed call for alternative interdisciplinary strategies that can invite a broader set of tools and research methods.

Only over the past decade, following the successful use of fluid analogy descriptions in production-inventory systems (Schwartz *et al.*, 2006), the Rivera *et al.*, 2007 article introduced the system identification and control engineering framework as a viable and promising approach for the design of novel adaptive interventions in the

context of behavioral health. In the article, it was pointed out that tools acquired from dynamical systems modeling and control have "the potential to significantly inform the analysis, design, and implementation of adaptive interventions, leading to improved adherence, better management of limited resources, a reduction of negative effects, and overall more effective interventions." Admittedly, over the last decade, this proposition has gained a sizable attraction from the behavioral science and medicine society, and has led to the development of new dynamical models of behavior change, which in turn inspired a family of adaptive intervention designs using control systems engineering ideas, each targeting a specific behavioral health challenge.

Toward addressing obesity and meeting a balanced dietary and physical activity lifestyle, similar to Schwartz *et al.*, Navarro-Barrientos *et al.*, 2011 proposed fluid analogy as a suitable tool for mechanistic modeling of complex dynamical systems, leading to the proposal of a dynamic behavior change model based on the Theory of Planned Behavior (TPB) (Ajzen, 1991). This TPB model was then integrated with a three-compartment, energy balance model developed in Hall and Jordan, 2008; Chow and Hall, 2008; Hall, 2009, 2010 and validated by the Minnesota Semi-Starvation Experiment (Keys *et al.*, 1950) to establish predictions of human weight gain/loss that are simultaneously driven by human psychology and physiology. Using the integrated behavioral-energy balance model, a classical feedback control system design was used to produce simulations illustrating optimized assignments of various intervention dosages over time. Moreover, the work of Dong *et al.*, 2012, 2013, 2014; Savage *et al.*, 2014 presented a receding horizon control strategy and a more elaborate intervention modeling that includes self-regulation and intervention delivery dynamics for designing improved GWG interventions and regulating infant birth weight; the potential in the proposed intervention strategies were illustrated by evaluating sim-

ulations of hypothetical human subjects. More recently, these efforts were followed by Guo *et al.*, 2016, 2017, 2018; Freigoun *et al.*, 2018; Pauley *et al.*, 2018; Symons Downs *et al.*, 2018 in which models of real human participants are estimated and validated from experimental data; these works have contributed in both modeling and improved intervention design for managing GWG and regulating infant birth weight.

Furthermore, alternative to traditional *static* (i.e., steady-state) modeling methods and "one-size-fits-all" strategies for smoking cessation, contributions from Timms *et al.*, 2013, 2014d,a,c,b presented an alternative dynamic modeling and control engineering-based intervention design using a mediational understanding of smoking cessation dynamics. An additional novel application of system identification and control was in the domain of fibromyalgia pain management (Deshpande *et al.*, 2011, 2012, 2014b,a). Indeed, the use of behavioral (or integrated behavioral-energy balance) models in cited literature has delivered an entirely novel conceptual paradigm for the design of personalized and perpetually adaptive behavioral health interventions. Without a doubt, this was realized by inviting system identification and control engineering principles to the design of behavioral health interventions that seek to improve and sustain adherence to recommended dietary and physical activity guidelines. This appeal to such a design is further magnified by the explosion of advanced mobile health (mHealth) sensor technologies (Nilsen *et al.*, 2017) that not only offer a cost-effective deployment of adaptive interventions at the largest of scales, but also facilitate the evaluation and improvement of existing behavioral theory models by leveraging intensive longitudinal data (Riley *et al.*, 2011).

In light of the discussed challenges, this dissertation is primarily motivated by the need to use real-life experimental studies to advance the system identification, control, and behavioral science research, including the development of new practical approaches and algorithms in these domains. To achieve this, two recent novel

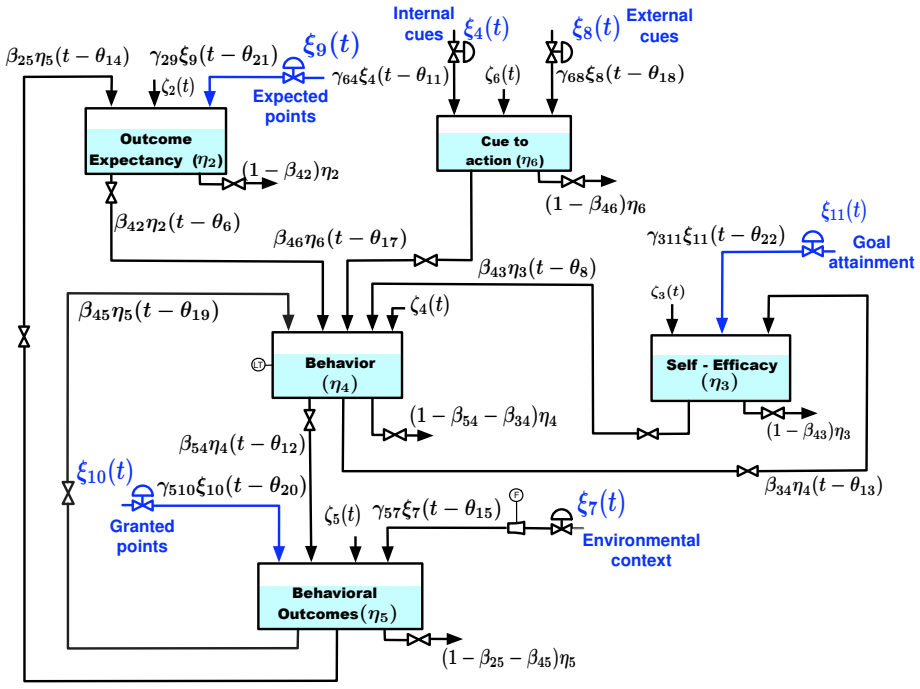

Figure 1.1: A Simplified Dynamical Systems Model of Social Cognitive Theory (Freigoun *et al.*, 2017; Hekler *et al.*, 2018).

open-loop intervention studies, *Healthy Mom Zone* (Symons Downs *et al.*, 2018), and *Just Walk* (Hekler *et al.*, 2018), are utilized in developing, estimating, and validating behavioral and energy balance models; an established participant-validated model will serve as basis for closed-loop intervention design. More specifically, this dissertation aims to answer calls from two 2014 papers: Savage *et al.* and Martín *et al.*. In the former, authors called for the need to utilize experimental data from a real-life intervention (such as *Healthy Mom Zone*) to improve, estimate, and validate their proposed maternal-fetal intervention model. As a result, chapter 2 is entirely devoted for the development of a dynamical systems model of intrauterine fetal growth, validated by data from *Healthy Mom Zone*.

The rest of this dissertation efforts are motivated by calls from Martín *et al.*, 2014, 2016a for estimating and validating the semiphysical model of Social Cognitive Theory (Bandura, 1986) in Figure 1.1 using data from real human subjects, and ultimately

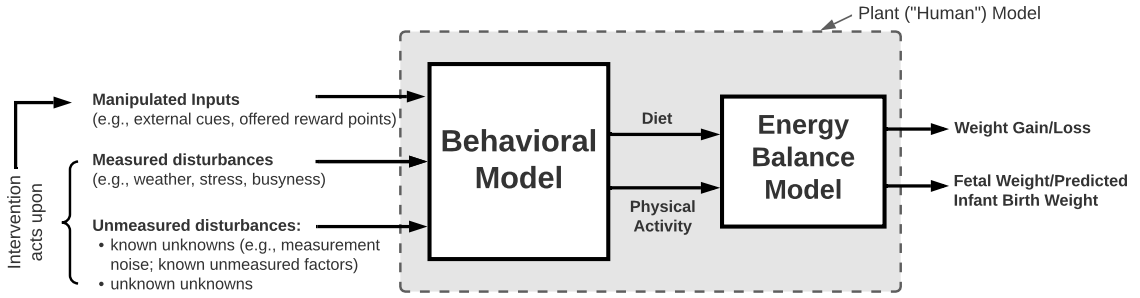

Figure 1.2: A Conceptual Block Diagram for an Integrated, Open-loop Behavioral Intervention Model for Managing GWG (and Fetal Growth) And/Or Obesity.

utilize these models to further support their hypothetical simulations of closed-loop physical activity interventions. In this work, these objectives are approached in the context of *Just Walk*, a more detailed description of which is provided in Chapter 3.

1.2 Modeling Behavioral Interventions

In the design of adaptive behavioral health interventions, depending on the target outcome (e.g., adherence to a prescribed level of daily physical activity, healthy eating, weight gain/loss, healthy infant birth weight), a model-based approach must utilize *a priori* knowledge of the underlying psychological and/or physiological mechanisms driving the particular outcomes of interest. For example, an intervention that aims to use a human psychology approach in order to achieve and maintain a prescribed level of physical activity will ultimately need to rely on a particular cognitive basis for predicting individual responses to various external cues and other contextual factors that influence behavior. A behavioral theory such as *Social Cognitive Theory* (SCT) can indeed lend a strong theoretical basis for establishing dynamical systems models that suit this purpose. Furthermore, if the intervention simultaneously targets weight gain/loss and promoting consistent physical activity, then it is natural to include an understanding for metabolic and other physiological mechanisms alongside the behavioral model on some level; this is also feasible by using an energy balance ap-

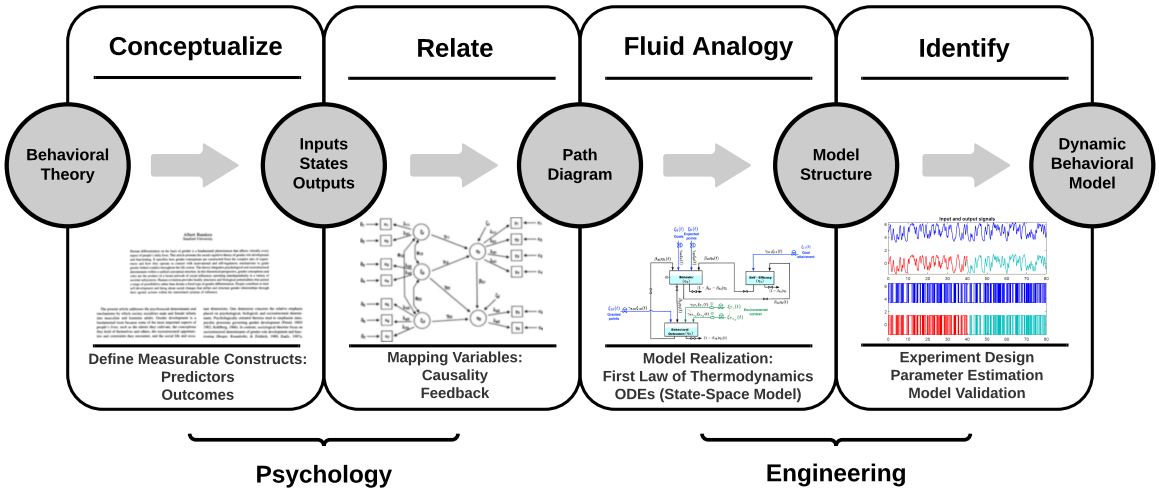

Figure 1.3: A Depiction of an Interdisciplinary Approach for Using Behavioral Theories to Estimate Dynamical Models Using Path Diagrams and Fluid Analogy.

proach. A block-diagram depiction of an integrated behavioral-energy balance model (open-loop) is presented in Figure 1.2.

Typically, energy balance models can be developed by applying the laws of thermodynamics, as well as the incorporation of empirical functions for some of the established variables. This will be visited with more depth in the context of developing a dynamical systems model of intrauterine fetal growth in Chapter 2. On the other hand, the modeling of behavioral theories such as SCT (Martín, 2016) or TPB (Guo, 2018) has proven to be a more challenging involved task in many respects. Figure 1.3 features a pathway from using behavioral theories to obtain control-oriented, dynamical models useful for adaptive intervention design. First, a behavioral theory in its abstract form is adopted, followed by harnessing all constructs (e.g., *self-efficacy*, *subjective norm*) originating from that theory. In the case of modeling and TPB, recent efforts (Navarro-Barrientos *et al.*, 2011; Dong, 2014; Martín, 2016; Guo, 2018) have collaborated in multidisciplinary research teams that included experts from the fields of psychology and behavioral medicine. Second, the considered behavioral theory is used to construct a path diagram characterizing the causal relations among the

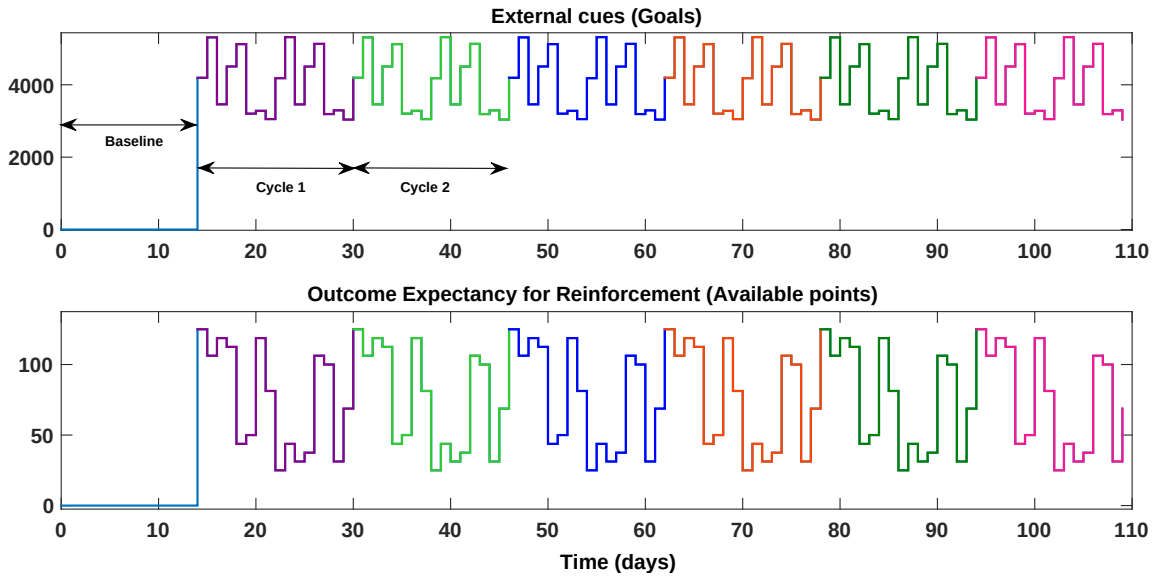

Figure 1.4: A Time-domain Visualization of the *Goals* and *Expected Points* Input Signals (See Figure 1.1) Using a "Zippered" Spectra Multisine Design for a Representative *Just Walk* Participant. *Goals* Are Measured in Steps/Day; *Expected Points* Are Measured over an Arbitrarily Constructed Point Scale (with 2,500 Points Redeeming a $5 Gift Card).

listed constructs (e.g., mediation, moderation, feedback) with the ultimate goal of describing a pathway from each measurable predictor (system input) to a measurable outcome (system output).

Next, the established path diagram can be used to construct a fluid analogy model (similar to Figure 1.1), which is most valuable for introducing the *dynamical* nature of all causalities framed by the obtained path diagram. For example, Figure 1.1 presents a *fifth-order* SCT model comprised of five interacting subsystems, each characterizing *first-order* dynamics. Using this fluid analogy model, a system of ordinary linear differential equations (ODEs) can be derived using the conservation of mass principle (i.e., the First Law of Thermodynamics) and, subsequently, a state-space model. Note that, conceptually, the *assumed* first law of thermodynamics need not be observed following the determination of a *behavioral* state-space representation.

Finally, with a state-space representation at hand, a "patient-friendly" experiment

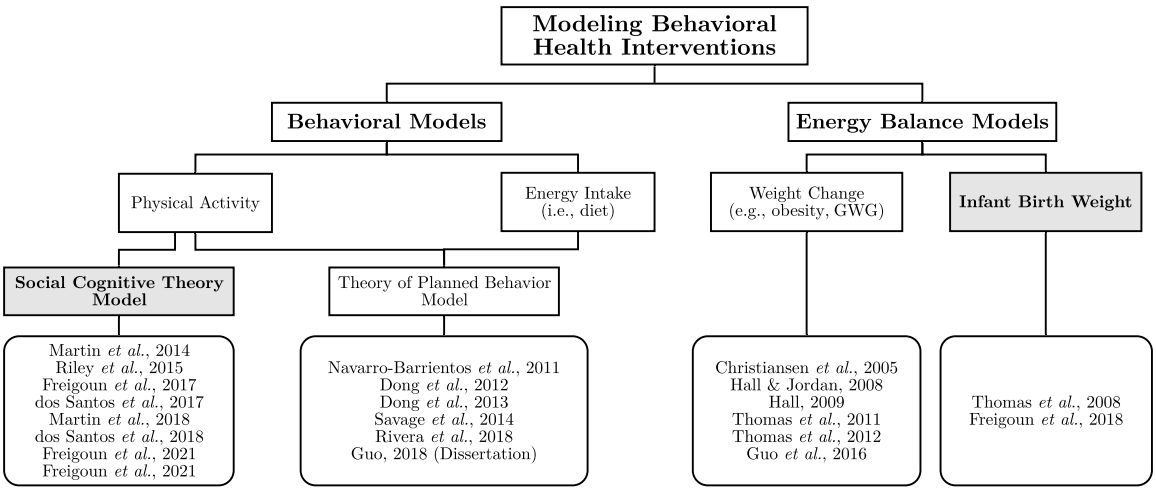

Figure 1.5: Scope (Shaded in Gray) of Modeling Work in This Dissertation with Respect to Selected Relevant Literature.

design based on system identification principles may be in order. In *Just Walk*, as will be further described in Chapter 3, the input signal design used sinusoidal excitations for the two input-output channels (*goals* and *expected points*) of the SCT model in Figure 1.1. The designed multisine perturbations feature a "zippered" spectra design for the purposes of generating time-domain signals that are orthogonal in frequency (Freigoun *et al.*, 2017). A time-domain visualization of the *goals* and *expected points* input signals using a "zippered" spectra multisine design for a representative *Just Walk* participant is shown in Figure 1.4. Finally, with a produced set of measured input-output data obtained from the identification experiment, parameter estimation and model validation follow; this is the necessary, final step that will be explored in depth for the SCT model in Figure 1.1 in the context of *Just Walk*. Some of the key contributions of this work is concerned with SCT model estimation and validation.

To further clarify the scope of this work, Figure 1.5 highlights areas of specific contributions in this application domain with respect to selected recently-produced body of work. The modeling of intrauterine fetal growth presented in Chapter 2 follows a *white-box* identification strategy, while the identification of participant-validated

SCT models using *Just Walk* relies on both, *black-* and *grey-box* estimation techniques. Since this terminology will be intensively used in the description of this work, the following introduction for the various types of dynamical systems models is borrowed from mainstream system identification (Ljung, 1994; Lindskog and Ljung, 1995; Lindskog, 1996; Ljung, 1999):

- *White-box (physically parameterized) models.* These are typically models of physical or biological systems that are derived from first-principles modeling (i.e., using laws of thermodynamics, motion, etc.), and incorporate all insights and knowns about the system behavior, usually in state-space form. They often contain known and unknown parameters; parameters have a physical significance of their own (i.e., meaning, units, etc.), such as unknown physical constants that are estimated from experimental data. While it is certainly possible to utilize these models in control system design (in fact are preferred if available), these are typically general-purpose, *theoretical* formulations that result from laborious modeling aimed at developing a better understanding and intuition for certain "truths" about the system in question.

- *Black-box models.* This type of models are data-driven, usually in the form of linear differential or difference equations. Neither the structure nor the parameters of a black-box model necessarily have physical significance. With the objective of fitting observed data to reproduce the past behavior of the system as accurate as possible, black-box models are typically easy to produce and used for *prediction/simulation* and/or *control* purposes. These types of models may also be estimated and studied in a preliminary step preceding structured modeling (Ljung, 1994).

- *Gray-box (semiphysical) models.* As the name implies, these models fall between

white- and black-box models, where prior knowledge about the concerned system is used in building a parsimonious model structure that is (i) intuitively descriptive of the system's components, and (ii) predictive of the system's dynamical behavior. However, building this structure is not carried out to the extent that a formal physically parametrized model is constructed since parameters may not have a direct physical interpretation. Gray-box system identification is particularly attractive in modeling and understanding causality in nonphysical systems and between abstract constructs. Indeed, in lieu of white-box models, gray-box models are also often employed in practice as "pragmatic" models (Ljung, 1999) when it is necessary or more practical to simplify the modeling of complex physical systems. Semiphysical models typically have the potential for serving all modeling purposes: *theoretical*, *control*, and *simulation/prediction*.

Upon the conclusion of all identification efforts, the established dynamical models can be used for control system design. Ultimately, a model-based intervention design that uses control systems engineering offer the ability to optimize intervention dosages according to individual characteristics (e.g., goal-driven, weekend warrior) and the ever-changing contextual factors (e.g., busyness, weather), while simultaneously respect economical and a host of other constraints.

1.3 Research Goals

At the highest level, the goal of this dissertation is to utilize experimental data drawn from recent novel behavioral interventions to provide a data-driven basis for supporting a number of recent works in this application domain. Almost at the same level, a main objective is to develop practical methods and glean insights that can prove useful for the future researcher or user interested in translating behavioral theories into dynamical models amenable to system identification and control system

design. More specifically, the goals of this dissertation include developing and validating a fetal model using *Healthy Mom Zone*, estimating and validating semiphysical Social Cognitive Theory models using *Just Walk*, and the evaluation of closed-loop simulations of participant-validated models using Hybrid Model Predictive Control.

1.3.1 Developing and Validating a Fetal Model Using Healthy Mom Zone

The underlying mechanisms for how maternal perinatal obesity and intrauterine environment influence fetal development are not well understood and thus require further understanding. In this dissertation, the goal is to use first principles modeling (i.e., energy balance and entropy concepts) for developing a comprehensive dynamical systems model for fetal growth that illustrates how maternal factors (energy intake and physical activity) influence fetal weight and related components (fat mass, fat-free mass, and placental volume) over time. Using *Healthy Mom Zone* (HMZ), a novel intervention for managing gestational weight gain in obese/overweight women, a more specific goal is to estimate and validate the developed fetoplacental model given intensive measurements of fetal weight and placental volume obtained from sonographic imaging technology. Ultimately, we aim to deliver a parsimonious system of equations that can reliably predict fetal weight gain and birth weight based on a sensible number of assessments, with a proven ability to inform clinical care recommendations and show how adaptive, HMZ-like interventions can influence fetal growth and birth outcomes.

1.3.2 Estimating and Validating Social Cognitive Theory Models Using Just Walk

Following the delivery of a more detailed description of *Just Walk*, an intensively adaptive physical activity intervention that has been designed on the basis of system identification and control engineering principles, a second major goal of this disserta-

tion is to further establish the viability of dynamical systems framework in capturing the key factors that influence and predict behavior change over time. To accomplish this, we aim to develop an unconventional ARX estimation-validation procedure that better suits *Just Walk* and seeks to balance predictive ability over validation data segments with overall goodness of fit. Black-box models (i.e., ARX models) come with the value of providing important clues to individual participant characteristics that influence physical activity; these insights will prove to be critical in building semiphysical models lending their theoretical basis from Social Cognitive Theory.

Further, while the semiphysical SCT model in Figure 1.1 can rise with favorable statistical properties compared to its fully-parametrized black-box counterpart, solver initialization of classical methods and structural identifiability often pose a challenge to the user seeking satisfactory results. Thus, a crucial step in this work is to apply a judicious model reduction thinking to propose a lower-complexity version of the SCT model in Figure 1.1; one goal is to show that this could be approached from both a conceptual and data-driven perspectives. Moreover, By assuming distinct poles and zero-order hold intersample behavior of the underlying system (similar to *Just Walk*), an important technical goal of this dissertation is show that the typical grey-box constrained optimization problem can be formulated into an easier one by solving linearly-constrained eigenvalue problems. Here, the proposed strategy is to follow the trend of existing literature and develop a formulation that relies on a consistent discrete-time black-box model (e.g., N4SID) to solve for a structured, continuous-time one in the absence of prior knowledge. Following the development of this grey-box identification algorithm, the goal is to establish semiphysical, participant-validated models of Social Cognitive Theory.

1.3.3 Closed-loop Simulations of Participant-Validated Models Using Hybrid Model Predictive Control

A final goal of this work is to produce and evaluate closed-loop intervention simulations of a participant-validated semiphysical SCT model from the previous section using published data from *Just Walk*. Consistent with real-world requirements, including the need for hybrid decision rules policies that factor in logical, physical, and financial constraints, it is only natural that considered closed-loop intervention designs in this dissertation would follow a Hybrid Model Predictive Control formulation. Since it is conceivable that clinical or psychological considerations may relax or require the gradual change of behavior, it is desired to examine simulations that feature single-phase as well as multi-phase interventions.

1.4 Contributions of the Dissertation

As highlighted earlier, the contributions from this work can be viewed under two major themes: First, harnessing experimental data to deliver an intrauterine fetal growth model relying on first principles. Second, the proven advancements in the domain of designing adaptive physical activity behavioral health interventions using Social Cognitive Theory models. In terms of the former part, the contributions of this dissertation are summarized as follows:

1. *A single-output energy balance model.* Building from the model in (Thomas *et al.*, 2008), the proposed energy balance model in this work features a single, easy-to-measure output (total fetal weight). In addition to grounding a better theoretical understanding of external factors and pre-existing conditions directly influence intrauterine fetal weight growth, this reformulation highlights less expensive and invasive requirements for estimating individualized model

parameters; measurements of total fetal weight can be far more reliable than measurements of body composition (more so in the first trimester; Bernstein and Catalano, 1991).

2. *Application of the Second Law of Thermodynamics.* The fetal energy balance model in this work provides a succinct, well-established accounting for the impact of entropy on fetal growth. Despite that the idea of entropy of new tissue formation has originated in Christiansen *et al.*, 2005, the work in Christiansen *et al.*, 2005 features an obesity model and the formulation cannot be directly re-purposed for quantifying fetal growth dynamics. Part of the contribution in this work is to merge efforts from Christiansen *et al.* and Thomas *et al.*: to produce a more rigorous and complete reformulation of fetal growth.

3. *Use of HMZ study data.* Utilizing data from the HMZ study (Symons Downs *et al.*, 2018), the developed fetal model presents a method for quantifying the impact of daily changes of physical activity on fetal growth. Moreover, using intensive, longitudinal participant data from HMZ, it is possible to estimate and validate the general first-principles fetal model structure developed in this work, as well as estimate a logistic profile of fetal fat mass accretion whose structure is supported by the literature.

4. *An improved placental volume model formulation.* As is discussed in Section 2.2.5, in this work, the curvature of the proposed placental volume model is more independently parameterized, which gives a more intuitive and easier model to estimate. This model also implicitly enforces the initial condition at conception; hence, for model estimation and simulation, the proposed model does not require a placental volume measurement for establishing an initial condition.

In terms of advancing intervention design using Social Cognitive Theory, the contributions of this work are:

1. *Black-box modeling of Just Walk.* Using input-output participant data, a library of ARX models were estimated and validated following a nontraditional, exhaustive modeling strategy. This contributions has proven useful as the same conceptual approach was applied in estimating and validating a host of MoliZoft models in dos Santos *et al.*, 2017, 2018, as well as in harnessing models that help improve a cocoa biofertilization process in Gallino *et al.*, 2018. Results and insights from this work has been cited in a relevant application setting that explore "modeling human-in-the-loop behavior and interactions with HVAC systems" (Kane, 2018).

2. *Semiphysical modeling of Just Walk..* In a necessary step mandated by the limitations of *Just Walk*, this work has introduced a further simplified, third-order semiphysical model of Social Cognitive Theory that is supported by experimental data. This further simplified model structure aims to capture behavior-change dynamics resulting from operant conditioning and self-efficacy. Above all, this dissertation claims to deliver the first estimated semiphysical models of Social Cognitive Theory using real input-output data drawn from an single-subject experiment that was designed based on system identification principles. The estimated models were validated using classical cross-validation tools from system identification and other disciplines.

3. *Identification framework for structured (grey-box) state-space models..* Inspired by challenges arising from *Just Walk*, a spectral decomposition identification algorithm was developed, along with its theoretical framework, and was intended to relieve the user from the burden of judicious solver initialization in

the absence of sufficient prior knowledge. Conditions relevant to the existence, uniqueness, and identifiability of *linear* grey-box structures under the proposed framework were provided. A numerical example illustrating the effectiveness of this formulation is provided. An extension to the main formulation that aims to accommodate *quadratic* structures was presented.

4. *Closed-loop simulations of participant-validated models.* Using HMPC control law together with a participant-validated SCT model from published *Just Walk* data, it is also claimed that the work of this dissertation is first to deliver closed-loop simulations of a data-driven participant model. By applying a simple controller reconfiguration, evaluated simulations in this work feature both single- and multi-phase intervention designs to provide more flexibility to the user.

1.5 Dissertation Outline

Following this introduction, the dissertation continues with Chapter 2 which introduces the *Healthy Mom Zone* study and utilizes it in the process of estimating and validating a quasi-linear parameter-varying fetoplacental model. The First and Second Laws of Thermodynamics are used to develop a dynamical systems, energy balance model of intrauterine fetal growth. Details of the estimation and validation steps are discussed, with simulations of multiple *Healthy Mom Zone* participants over the second and third trimesters presented.

Next, this dissertation moves to introduce the modeling and identification efforts using *Just Walk* in Chapter 3. Further details about the *Just Walk* intervention is provided, including the experimental design methodology. Further, results obtained from black-box modeling and estimation are presented, underscoring the amenability of the dynamical systems framework to capturing and predicting human behavior

change over time. Results from human subject models highlighting the idiosyncratic nature of human behavior are outlined.

In Chapter 4, a spectral decomposition identification formulation for structured state-space models is delivered along with its theoretical foundation. This general-purpose formulation is intended to address one of the long-existing grey-box identification challenges that presented a barrier to achieving some of the goals of this work. The developed algorithm is utilized in the process of estimating data-driven semi-physical models of Social Cognitive Theory. Results establishing a basis for model validation are also included.

Using a semiphysical model of Social Cognitive Theory developed in Chapter 4, Chapter 5 features an evaluation for the time-domain responses of a *Just Walk* participant. A brief overview on the general and application-specific HMPC formulations is presented. Further, the participant-validated SCT model is used to produce and evaluate HMPC-governed closed-loop simulations of a physical activity intervention. Both single- and multi-phase intervention designs are proposed and simulated.

Finally, the dissertation concludes with Chapter 6, providing a summary and some of the key conclusions stemming from this work. The chapter includes a brief discussion of some of the potential extensions to this work.

1.6 Publications

The following are lists of conference proceedings and journal papers highlighting published contributions over the course of this research.

1.6.1 Conference Proceedings

1. Freigoun, M. T., C. A. Martín, A. B. Magann, D. E. Rivera, S. S. Phatak, E. V. Korinek and E. B. Hekler, "System Identification of *Just Walk*: A Behavioral

mHealth Intervention for Promoting Physical Activity", in "American Control Conference (ACC), 2017", pp. 116-121 (2017). Chapter 3 is adapted from this publication.

2. dos Santos, P. L., M. T. Freigoun, D.E. Rivera, E.B. Hekler, C.A. Martín, R. Romano, T. Perdicoúlis and J. Ramos, "A MoliZoft System Identification Approach of the Just Walk Data", IFACPapersOnLine 50, 1, 12508-12513 (2017).

3. Freigoun, M. T., K. S. Tsakalis and G. B. Raupp, "A Spectral Decomposition Identification Algorithm for Structured State-Space Models: Estimating Semiphysical Models of Social Cognitive Theory", in "American Control Conference (ACC), 2021", pp. 2830-2835 (2021b). Chapter 4 is adapted from this manuscript.

4. Freigoun, M. T., K. S. Tsakalis and G. B. Raupp, "On the Identification of Social Cognitive Theory Models and Closed-loop Intervention Simulations Using Hybrid Model Predictive Control", Accepted for publication in Proceedings of the 19$^{\text{th}}$ IFAC Symposium on System Identification (2021a). Chapter 5 is adapted from this manuscript.

1.6.2 Journal Articles

1. Phatak, S. S., M. T. Freigoun, Martín, D. E. Rivera, E. V. Korinek, M. A. Adams, M. P. Buman, P. Klasnja and E. B. Hekler, "Modeling Individual Differences: A Case Study of the Application of System Identification for Personalizing a Physical Activity Intervention", *Journal of Biomedical Informatics* 79, 82-97 (2018).

2. Korinek, E. V., S. S. Phatak, C. A. Martín, M. T. Freigoun, D. E. Rivera, M. A. Adams, P. Klasnja, M. P. Buman and E. B. Hekler, "Adaptive Step Goals and

Rewards: A Longitudinal Growth Model of Daily Steps for a Smartphone-based Walking Intervention", *Journal of Behavioral Medicine* 41, 1, 74-86 (2018).

3. Hekler, E. B., D. E. Rivera, C. A. Martín, S. S. Phatak, M. T. Freigoun, E. Korinek, P. Klasnja, M. A. Adams and M. P. Buman, "Tutorial for Using Control Systems Engineering to Optimize Adaptive Mobile Health Interventions", *Journal of Medical Internet Researc*h 20, 6 (2018).

4. Freigoun, M. T., D. E. Rivera, P. Guo, E. E. Hohman, A. D. Gernand, D. Symons Downs and J. S. Savage, "A Dynamical Systems Model of Intrauterine Fetal Growth", *Mathematical and Computer Modelling of Dynamical Systems* 24, 6, 641-667 (2018). Chapter 2 is adapted from this publication.

5. dos Santos, P. L., M. T. Freigoun, C. A. Martín, D. E. Rivera, E. B. Hekler, R. A. Romano and T. P. Perdicoúlis, "System Identification of *Just Walk*: Using Matchable-Observable Linear Parametrizations", *IEEE Transactions on Control Systems and Technology* (2018).

Chapter 2

A DYNAMICAL SYSTEMS MODEL OF INTRAUTERINE FETAL GROWTH

2.1 Background

Prior work has described a conceptual framework for managing GWG among overweight/obese women (Symons Downs *et al.*, 2018) and for regulating infant birth weight (Savage *et al.*, 2014); this framework relies on methods from control systems engineering to develop decision policies that optimize the adaptation for participant response. The implementation of such a framework calls for developing advanced control systems which rely on dynamical models that are able to predict individualized responses to different intervention components and subsequently predict GWG, the intrauterine growth profile, and infant birth weight (Savage *et al.*, 2014; Thomas *et al.*, 2012; Guo *et al.*, 2016). In particular, one important end use of a dynamical systems model of intrauterine fetal growth is as the internal model in a model-based controller that accomplishes an optimized, adaptive intervention (Rivera *et al.*, 2007, 2017, 2018).

Energy balance for modeling weight and body composition change has been examined extensively, including among pregnant women (Hall, 2014; Thomas *et al.*, 2012). Modeling intrauterine growth has received some prior examination Thomas *et al.*, 2008; however, further modeling efforts are needed to better understand how prenatal status 'programs' fetal growth (Chandler-Laney and Bush, 2011; Catalano and Ehrenberg, 2006). To address this gap, we use intensive longitudinal data from *Healthy Mom Zone* (HMZ) (Symons Downs *et al.*, 2018), an ongoing trial, which is an individually-tailored, adaptive intervention to manage weight gain in overweight and obese preg-

nant women. While the model developed in this work extends from prior work (Thomas *et al.*, 2008; Christiansen *et al.*, 2005), it grounds a more complete theoretical understanding for how external maternal factors (e.g., daily energy intake and physical activity) influence fetal growth profiles.

In this chapter, we present parameter estimation and model validation results drawn from four representative HMZ participants. The final fetal energy balance model parameters are estimated by solving a nonlinear least squares optimization problem; the set of estimated model parameters is then used to generate simulations for model validation.

This chapter is organized as follows: Section 2.2 presents the underlying modeling assumptions and describes the derivation of the proposed fetal energy balance model. Section 2.3 features the optimization problem that accomplishes model parameter estimation from ultrasound measurements, followed by a presentation of the metrics and criteria used for model validation. Section 2.4 summarizes conclusions and future work.

2.2 Fetal Energy Balance Model

We begin by outlining important assumptions and simplifications leading to the final proposed fetal energy balance model. Next, building on insights from prior researchers (Christiansen *et al.*, 2005; Thomas *et al.*, 2008), we establish a first-principles energy balance model of fetal growth. Following the first law of thermodynamics, this fetal energy balance model applies the conservation of energy principle. Further, the presented derivation explicitly accounts for the energy loss due to new fetal tissue formation, as dictated by the second law of thermodynamics from which it follows that the conversion of energy requires energy. Finally, explicitly defined logistic growth functions are established to estimate the rate of fetal fat mass deposition

and the placental volume.

2.2.1 Initial Assumptions

The following initial assumptions and simplifications lead to the proposed fetal model (equation (2.21)):

1. Fetal body mass is divided into two main components: fat and fat-free tissues.
2. Fetal energy expenditure due to diet-induced thermogenesis is negligible.
3. Fetal physical activity in the womb is negligible.
4. The rate of fetal fat mass deposition is only regulated by the total fetal body mass (Christiansen *et al.*, 2005).
5. The contribution of daily maternal diet to fetal nutrition substantially exceeds additional nutrient supply originating from maternal body components.
6. Fetal energy imbalances are always positive and follow from the diet of a healthy, well-nourished mother.
7. The proportion of fetal body fat that contributes to expenditure is equal to that of fetal fat-free tissues.

Following the derivation of the fetal model in Sections 2.2.2 and 2.2.3, a discussion of the rationale for assumptions 6. and 7. is provided.

2.2.2 Energy Balance Equation

The basis for determining fetal growth is a daily energy balance based on the First Law of Thermodynamics that takes into account the metabolizable energy intake I_f (provided by the mother) and fetal energy expenditure $E_{ef}(t)$ to define a rate of

Table 2.1: Glossary of the Fetal Model Constants, Parameters, and Variables

Constants		
λ_{FM_f}	Energy stored per unit fetal fat mass	[kcal/kg]
λ_{FFM_f}	Energy stored per unit fetal fat-free mass	[kcal/kg]
μ	Daily energy expenditure per unit fetal body mass	[kcal/kg/d]
Parameters		
e_{FM_f}, e_{FFM_f}	Efficiencies of conversion of excess energy to new fat and fat-free tissues, respectively	[1]
α	Proportionality constant	[d/kcal/ml]
γ	Conversion coefficient	[ml^{-1}]
Variables		
t	Gestational age	[days]
$C_f(t)$	Daily energy accumulation in the fetus	[kcal]
$I_f(t)$	Daily fetal energy intake resulting from maternal energy intake	[kcal/d]
$E_{e_f}(t)$	Total fetal energy expenditure	[kcal/d]
$E_{Mf}(t)$	Energy required to maintain the fetus life	[kcal/d]
$E_{cf}(t)$	Energy required for the conversion of excess energy into new fetal tissue	[kcal/d]
$FM_f(t)$	Fetal fat mass	[kg]
$FFM_f(t)$	Fetal fat-free mass	[kg]
$W_f(t)$	Total fetal weight	[kg]
$f_r(W_f)$	Rate of fetal fat mass deposition	[1=$\text{kg}_{\Delta FM_f}/\text{kg}_{\Delta W_f}$]
$W_m(t)$	Total maternal weight	[kg]
$E_f(t)$	Total energy to build the fetal tissue up to day t	[kcal]
$E_{FM_f}(t)$	Total energy to build the fetal fat tissue up to day t	[kcal]
$E_{FFM_f}(t)$	Total energy to build the fetal fat-free tissue up to day t	[kcal]
$m(t)$	Maternal energy intake	[kcal/d]
$PA(t)$	Maternal physical activity	[kcal/d]
$P(t)$	Placental volume	[ml]
$g(t)$	Glycemic impact of intake	[1]
$K_f(t)$	Fetal gain coefficient from intake	[kg·d/kcal/ml]
$\tau_f(W_f)$	Time constant of fetal weight growth	[d]
$e_f(W_f)$	Overall efficiency of energy conversion	[1]

accumulation of the total fetal energy $C_f(t)$. Considering the fetus as the system of interest, we have

$$\begin{array}{ccccc} \text{Rate of Energy} & & \text{Energy} & & \text{Energy} \\ \text{Accumulation} & = & \text{Intake Per Day} & - & \text{Expenditure Per Day} \\ \frac{dC_f(t)}{dt} & = & I_f(t) & - & E_{e_f}(t) \end{array} \tag{2.1}$$

with

$$E_{e_f}(t) = E_{e_f} = E_{M_f}(t) + E_{c_f}(t) \tag{2.2}$$

accounting for fetal energy expenditure towards maintaining and sustaining life (E_{Mf}) and the energy required for the conversion of excess energy into new tissue (E_{cf}). A full glossary of model components is presented in Table 2.1.

Considering a two-compartment energy balance model (i.e., total body mass divided into fat and fat-free mass components), the positive rate of change of the total combustible fetal energy content, dC_f/dt, can also be calculated by accounting for changes of fetal body components Thomas *et al.*, 2008, giving

$$\frac{dC_f(t)}{dt} = \lambda_{FM_f}\frac{dFM_f(t)}{dt} + \lambda_{FFM_f}\frac{dFFM_f(t)}{dt} \tag{2.3}$$

which, in turn, when combined with (2.1), yields

$$\lambda_{FM_f}\frac{dFM_f(t)}{dt} + \lambda_{FFM_f}\frac{dFFM_f(t)}{dt} = I_f(t) - E_{ef}(t) \tag{2.4}$$

with $FM_f(t)$ and $FFM_f(t)$ denoting the total fetal fat and fat-free masses, respectively; λ_{FM_f} and λ_{FFM_f} are the energy densities of the fetal fat and fat-free components, respectively (i.e., energy content per unit fat/fat-free mass). As depicted in equation (2.4), both λ_{FM_f} and λ_{FFM_f} are assumed time-invariant.

Equation (2.4) presents the basic fetal energy balance result following directly from the first law of thermodynamics, as similarly highlighted in the Thomas *et al.*, 2008

model. However, equation (2.4) involves terms that need to be further defined, are difficult to measure experimentally, or expensive to track in an intervention setting. More specifically, in this work, profiles describing the evolution of the fetal body composition (FM and FFM), portion of maternal energy intake contributing to fetal nutrition, and the influence of maternal physical activity are all terms that are explored and expanded further from (2.4). Moreover, the expenditure term, $E_{ef}(t)$, requires estimates for the efficiency of energy conversion into new fetal fat and fat-free tissues (energy loss due to entropy); these efficiencies are difficult and expensive to measure experimentally. Furthermore, given current imaging technologies that build from well-studied sonographic methods to estimate total fetal weight, it is advantageous to reformulate the basic fetal energy balance equation shown as (2.4) in terms of the total fetal body mass (also referred to as total fetal weight, $W_f(t)$). In the following section, the primary aim is to establish a parsimonious fetal energy balance model that proves to overcome these challenges.

2.2.3 Efficiency of Energy Conversion & Energy Balance Reformulation

The goal of this section is to formulate equation (2.4) in terms of total fetal weight. To achieve this outcome, we built from concepts used to develop human obesity models by Christiansen *et al.*, 2005. The time-varying rate of fetal fat mass deposition (with respect to total fetal weight) is defined as follows:

$$f_r(W_f) \stackrel{\text{def}}{=} \lim_{\Delta W_f \to 0} \frac{\Delta FM_f}{\Delta W_f} = \frac{dFM_f}{dW_f} \tag{2.5}$$

which leads to the following expressions for the rate of change of FM_f and FFM_f in terms of total fetal weight W_f,

$$\frac{dFM_f}{dt} = \frac{dFM_f}{dW_f}\frac{dW_f}{dt} \qquad \triangleq f_r(W_f)\frac{dW_f}{dt} \tag{2.6a}$$

$$\frac{dFFM_f}{dt} \triangleq \frac{d}{dt}\left(W_f - FM_f\right) \qquad = \left[1 - f_r(W_f)\right]\frac{dW_f}{dt} \tag{2.6b}$$

With an explicitly defined $f_r(W_f)$, the components $FM_f(t)$ and $FFM_f(t)$ become explicit functions of the total fetal weight, $W_f(t)$. Using equations (2.3) and (2.6), we now have

$$\frac{dC_f}{dt} = \left(\lambda_{FM_f} f_r(W_f) + \lambda_{FFM_f}\left[1 - f_r(W_f)\right]\right)\frac{dW_f}{dt} \tag{2.7}$$

Second, we also define the efficiencies of new fetal tissue formation arising from the second law of thermodynamics as follows (Çengel and Boles, 2005):

$$\text{efficiency of fat mass deposition} = \quad e_{FM_f} \quad \stackrel{\text{def}}{=} \quad \lambda_{FM_f}\frac{dFM_f}{dE_{FM_f}} \tag{2.8a}$$

$$\text{efficiency of fat-free mass deposition} = \quad e_{FFM_f} \quad \stackrel{\text{def}}{=} \quad \lambda_{FFM_f}\frac{dFFM_f}{dE_{FFM_f}} \tag{2.8b}$$

where $dE_f \triangleq dE_{FMf} + dE_{FFMf}$ captures the total energy required to increase the total fetal body energy content by dC_f. The efficiencies in (2.8) provide a useful parametric representation for the energy loss due to new fetal fat and fat-free tissue formation, respectively. Thus, from equations (2.6) and (2.8) we have

$$\begin{aligned}\frac{dE_f}{dt} \triangleq \frac{dE_{FMf}}{dt} + \frac{dE_{FFMf}}{dt} &= \frac{\lambda_{FM_f}}{e_{FM_f}}\frac{dFM_f}{dt} + \frac{\lambda_{FFM_f}}{e_{FFM_f}}\frac{dFFM_f}{dt} \\ &= \left(\frac{\lambda_{FM_f} f_r(W_f)}{e_{FM_f}} + \frac{\lambda_{FFM_f}\left[1 - f_r(W_f)\right]}{e_{FFM_f}}\right)\frac{dW_f}{dt}\end{aligned} \tag{2.9}$$

As first realized by Christiansen *et al.*, 2005, the dynamic rate of change of E_f can be calculated by establishing the available energy for new fetal tissue deposition; i.e., the difference between the fetal energy intake and the energy expenditure required

for sustaining and maintaining life of existing fetal tissues, thus

$$\frac{dE_f}{dt} \triangleq I_f - E_{Mf} \tag{2.10}$$

Combining (2.9) and (2.10) gives

$$\frac{dW_f}{dt} = \frac{dW_f}{dE_f}\frac{dE_f}{dt} = \frac{I_f - E_{Mf}}{\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f}\left[1 - f_r(W_f)\right]/e_{FFM_f}} \tag{2.11}$$

which, when substituted into (2.7), gives

$$\frac{dC_f}{dt} = \underbrace{\left(\frac{\lambda_{FM_f} f_r(W_f) + \lambda_{FFM_f}\left[1 - f_r(W_f)\right]}{\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f}\left[1 - f_r(W_f)\right]/e_{FFM_f}}\right)}_{e_f(W_f)} \left(I_f - E_{Mf}\right) \tag{2.12}$$

where now the ratio between dC_f/dt and $\left(I_f - E_{Mf}\right)$ represents the overall *time-varying* thermodynamic efficiency of energy conversion into new fetal tissue, $0 \leq e_f(W_f) \leq 1$; this was first similarly established by Christiansen *et al.*, 2005, however, with a constant f_r assumed. Inserting equation (2.2) into equation (2.1) and contrasting with (2.12) provides an accounting method for the energy loss due to new tissue formation

$$E_{c_f}(t) = \left(1 - e_f(W_f)\right)\left(I_f(t) - E_{M_f}(t)\right) \tag{2.13}$$

with $e_f(W_f)$ per equation (2.12).

To establish the $I_f(t)$ term in (2.12), it is known that the fetal energy intake through the placenta (whose volume is denoted by $P(t)$) originates mainly from two nutritional sources: maternal diet, $m(t)$, and maternal body components (e.g., muscles, fats, bones, etc.) Barker, 2008, hence giving

$$\hat{I}_f(t) = \gamma(t)\left[g(t)m(t) + \alpha_W(t)W_m(t)\right]P(t) \tag{2.14}$$

where $W_m(t)$ is daily total maternal weight; $\alpha_W(t)$ is a function that captures the daily fraction of maternal body mass directly contributing to fetal nutrition ([=]

kcal/kg/d). $g(t)$ denotes the daily glycemic impact of intake (ranges from 0 to 1); $\gamma(t)$ is a conversion coefficient that is associated with maternal physical activity, as postulated in equation (2.22) (and discussed later in the chapter). However, for the case of a healthy, non-fasting and well-nourished mother, it may be accepted to assume that the basic nutritional needs for fetal growth can be met by daily maternal diet alone (Langhoff-Roos *et al.*, 1987; Cetin *et al.*, 2009). Hence, it is assumed that $g(t)m(t) \gg \alpha_W(t)W_m(t) \quad \forall t$, giving (identical to Thomas *et al.*, 2008)

$$I_f(t) = \gamma(t)m(t)g(t)P(t) \tag{2.15}$$

The fetal energy expenditure term in (2.12) ($E_{M_f}(t)$) can be considered, for simplicity, as a direct proportion of total fetal body mass Thomas *et al.*, 2008:

$$E_{Mf}(t) = \mu\left[FM_f(t) + FFM_f(t)\right] \triangleq \mu W_f(t) \tag{2.16}$$

where μ is the daily energy expenditure per unit fetal body mass. Hence, from (2.7) and (2.12) we have

$$\left(\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f}\left[1 - f_r(W_f)\right]/e_{FFM_f}\right)\frac{dW_f(t)}{dt} = I_f(t) - E_{Mf}(t) \tag{2.17}$$

Applying equations (2.15) and (2.16) gives

$$\left(\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f}\left[1 - f_r(W_f)\right]/e_{FFM_f}\right)\frac{dW_f(t)}{dt} = \gamma(t)m(t)g(t)P(t) - \mu W_f(t) \tag{2.18}$$

Furthermore, dividing equation (2.18) by μ and defining

$$K_f(t) = \frac{\gamma(t)g(t)}{\mu} \tag{2.19}$$

$$\tau_f(W_f) = \frac{\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f}\left[1 - f_r(W_f)\right]/e_{FFM_f}}{\mu} \tag{2.20}$$

yields a final fetal energy balance equation in terms of the total fetal weight:

$$\tau_f(W_f)\frac{dW_f(t)}{dt} + W_f(t) = K_f(t)m(t)P(t) \tag{2.21}$$

Equation (2.21) features an intrauterine fetal weight growth model that conforms with the description of a first-order quasi Linear Parameter-Varying (quasi-LPV) system whose scheduling variable is the output, i.e., the total fetal weight, $W_f(t)$. In equation (2.21), the growth parameter τ_f is the time constant (Ogunnaike and Ray, 1994) which characterizes the speed of response. The parameters $\gamma(t)$ and $g(t)$ appearing in equation (2.19) are discussed in the explanation of equations (2.22) and (2.30), respectively.

In addition to achieving the goal of reformulating equation (2.4) in terms of a single, measurable output variable (i.e., the total fetal weight), equation (2.21) features an intuitive, well-understood first-order dynamical systems model structure that is more amenable to system identification and control. A further outcome following from the development of (2.21) is that estimates for e_{FM_f} and e_{FFM_f} can be determined directly from ultrasound measurements.

Following the development of the model in equation (2.21) we make the following remarks:

- Given that the exact mechanism governing the influence of maternal physical activity on fetal weight is yet to become sufficiently understood, we follow Thomas *et al.*, 2008 in assuming that maternal physical activity influences the placenta function, and thereby influences the fetus' nutrition. This is further established in Clapp, 2003 from which it is known that the effect of maternal exercise on fetal growth depends on numerous factors such as type, frequency, intensity, and the time point in pregnancy when the exercise is performed. Hence, for simplicity, we assume that, over a baseline, maternal physical activity is proportional to placental function, which is captured via the $\gamma(t)$ parameter in (2.21); i.e.,

$$\gamma(t) = \alpha PA(t) + \beta \tag{2.22}$$

where $PA(t)$ denotes the daily maternal physical activity, α is the proportionality constant, and β is the established baseline. Following the literature review presented by Thomas *et al.*, 2008, we further assume that $\alpha \leq 0$ and $\beta \geq \gamma > 0 \quad \forall\, t \geq 0$.

- It follows from assumption 6. that $dW_f/dt \geq 0$ during gestation; thus, from equation (2.21) we have

$$K_f(t) = \frac{\gamma(t)g(t)}{\mu} \geq \frac{W_f(t)}{m(t)P(t)} \qquad \forall t \ \text{ during gestation} \tag{2.23}$$

providing one important criterion for model validation. Additionally, the inequality in (2.23) can serve as an approximate (yet useful) diagnostic tool indicating rate of fetal growth (as will be discussed later in Figure 2.11).

- Kennaugh and Hay Jr, 1987 reported estimates where μ_{FM_f} and μ_{FFM_f} need not be averaged into a single proportion of energy expenditure (energy requirement); in which case, contrary to assumption 7., if $\mu_{FM_f} \neq \mu_{FFM_f}$, it can be shown that equation (2.21) becomes

$$\tau_f'(W_f)\frac{dW_f(t)}{dt} + W_f(t) = K_f'(t)m(t)P(t) + \left(\frac{\mu_{FFM_f} - \mu_{FM_f}}{\mu_{FFM_f}}\right)\int_0^{W_f(t)} f_r(W_f)dW_f \tag{2.24}$$

with

$$K_f'(t) = \frac{\gamma(t)g(t)}{\mu_{FFM_f}}, \quad \tau_f'(W_f) = \frac{\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f}\left[1 - f_r(W_f)\right]/e_{FFM_f}}{\mu_{FFM_f}}$$

where μ_{FM_f} and μ_{FFM_f} are the proportions of energy expenditure corresponding to maintaining and sustaining life of fetal fat and fat-free tissues, respectively. Nonetheless, given the desire for a parsimonious model, we continue to assume that $\mu_{FM_f} = \mu_{FFM_f} = \mu$, where equation (2.21) applies.

- The model parameters $\gamma(t)$, e_{FM_f}, and e_{FFM_f} are assumed to vary on an individual level. According to the formulation of (2.21), these parameters may capture between- and/or within-group differences (e.g., genetic variations (Noblet *et al.*, 1999), exercising vs. non-exercising).

2.2.4 Rate of Fetal Fat Mass Deposition

From the previous discussion, the importance of understanding the rate of fetal fat deposition, $f_r(W_f)$, as a key variable to attaining a predictive fetal energy balance model is now clear. From data presented and analyzed in a fetal body composition study by Demerath *et al.*, 2016, good *a priori* knowledge is now available to establish the dependence of f_r on $W_f(t)$. Literature also strongly suggests that the accretion of fetal fat starts accumulating after 26-30 weeks gestation (Schwartz and Galan, 2003; Thomas *et al.*, 2008); to this effect, the following piecewise modified logistic equation can be considered

$$FM_f(t) = FM_f(W_f(t)) = \begin{cases} \dfrac{c_{f_r}}{1 + e^{-a_{f_r}\left[W_f(t) - b_{f_r}\right]}} + C & t \geq t_0 \\ 0 & t < t_0 \end{cases} \tag{2.25}$$

with identifiable parameters a_{f_r}, b_{f_r}, c_{f_r}, and initial time, t_0, estimated as described in Section 2.3; C is a constant. When $FM_f(W_{f_0}) = 0$ at t_0, we get

$$C = -\frac{c_{f_r}}{1 + e^{-a_{f_r}\left(W_{f_0} - b_{f_r}\right)}}$$

where W_{f_0} is the initial weight at the initial time t_0. From equations (2.5) and (2.25), f_r is now an explicitly defined function; namely,

$$f_r(W_f) = \begin{cases} a_{f_r} c_{f_r} \dfrac{e^{-a_{f_r}\left(W_f(t) - b_{f_r}\right)}}{\left[1 + e^{-a_{f_r}\left(W_f(t) - b_{f_r}\right)}\right]^2} & t \geq t_0 \\ 0 & t < t_0 \end{cases} \tag{2.26}$$

For simplicity, in this work we will assume that $t_0 = 0$ (or $W_{f_0} = 0$). Finally, given a well-defined f_r (equal to (2.26) or otherwise), estimations of the fetal body components readily follow from equation (2.6); namely,

$$FM_f(t) = \int_0^{W_f(t)} f_r(W_f) dW_f \tag{2.27}$$

$$FFM_f(t) = W_f(t) - \int_0^{W_f(t)} f_r(W_f) dW_f \tag{2.28}$$

2.2.5 Placental Volume Growth Model

Following Thomas *et al.* (2008) Thomas *et al.*, 2008; Azpurua *et al.*, 2010, we consider the placental volume $P(t)$ as the most suitable variable to characterize placental growth. There is a substantial literature where placental development and growth profiles are presented and characterized throughout gestation for humans (Pitkin, 1976; Thompson *et al.*, 2007; Wallace *et al.*, 2013) and animals (Mu *et al.*, 2008). The placenta grows in three phases: first, a 'lag' phase in which cells begin to form; second, an exponential growth phase where cells continue to form and rapidly divide; and finally, due to space restrictions, a deceleration in the growth rate is expected in the final weeks towards birth. These three growth phases are adequately captured with a logistic growth function. Figure 2.1 features a standard logistic growth profile where these three phases are depicted.

Similar to equation (2.25), the 'modified' logistic function is considered

$$P(t) = c_P \left[\frac{1}{1 + e^{-a_P(t-b_P)}} - \frac{1}{1 + e^{a_P b_P}} \right] \qquad \forall\ t \geq 0 \tag{2.29}$$

with $P(0) = 0$ and $\lim\limits_{a_P b_P \to \infty} \lim\limits_{t \to \infty} P(t) = c_P \lim\limits_{a_P b_P \to \infty} \left(\frac{e^{a_P b_P}}{1 + e^{a_P b_P}} \right) = c_P$, where a_P, b_P, and c_P (c_P is the 'ultimate' carrying capacity) are identifiable model parameters from the estimation procedure described in Section 2.3.7; it follows from Figure 2.1 that

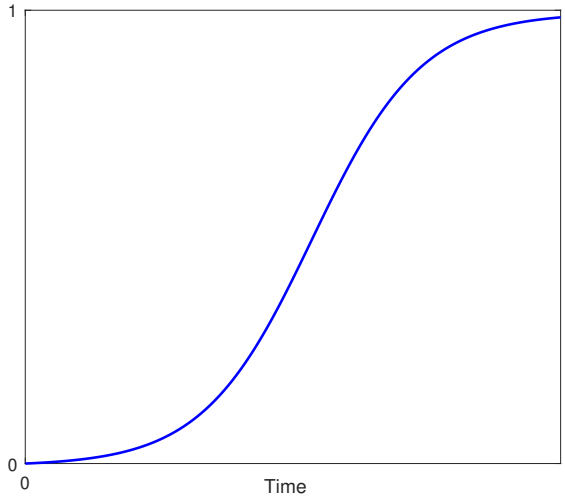

Figure 2.1: Representative Placental Volume Growth Profile.

$a_P, b_P, c_P \geq 0$. The algebraic model in equation (2.29) differs from the placental volume in Thomas *et al.*, 2008 in that its curvature features are more independently parameterized: the parameter a_P assigns the rate of growth, b_P assigns the inflection point, and c_P assigns the ultimate carrying capacity or the scale of the profile (note $c_P = 1$ in Figure 2.1 for illustration). In Thomas *et al.*, 2008, the proposed model does not apply when the initial condition is $P(0) = 0$, and hence, requires an additional estimated ultrasound measurement of EPV. Moreover, the parameters (including the initial condition) of the model in Thomas *et al.*, 2008 play simultaneous role in determining its final curvature features, which makes it less intuitive.

It has been reported that the size and growth rates of the placenta are associated with physical activity (Thomas *et al.*, 2008; Clapp, 2003) and additional genetic factors (Regnault *et al.*, 2001). In the presence of more intensive ultrasound measurements, the carrying capacity parameter c_P can be further investigated such that moderations of placenta size over time by physical activity or genetic differences are more understood; this also applies to the growth rate a_P and mid-point b_P parameters. In our parameter estimation, we assume constant parameters a_P, b_P, and c_P

such that averaged, fixed-effects are captured.

2.3 Parameter Estimation and Model Validation

The *Healthy Mom Zone* (HMZ) study (Symons Downs *et al.*, 2018) is an individually-tailored, adaptive behavioral intervention for managing weight in pregnant women with overweight and obesity. The target sample is 30 pregnant women who are randomized to either the intervention or control group from approximately 8 to 36 weeks gestation. Study measures including weight, physical activity, and energy intake. are obtained at baseline, throughout the course of the intervention (e.g., daily, weekly, or monthly), and at follow-up. The detailed intervention protocols that includes eligibility, recruitment, intervention description, dosages, and measurement schedule have been published elsewhere (Symons Downs *et al.*, 2018). In addition, an ancillary project provides six ultrasound measures used to estimate fetal weight, placental volume, and fetal body composition. In this section, further discussion of each estimated measurement is presented; four representative completed participants ($n = 4$; three overweight, one obese; mean age=30.3 years, two intervention, two control) are considered.

2.3.1 Estimated Fetal Weight

An estimated fetal weight (EFW) can be drawn from ultrasounds when specific biomarkers are measured as displayed in the example Figure 2.2. Using these biomarkers, one of the best known and well-established correlations that can be applied is the Hadlock estimation. For our model estimation, we use a set of six EFW measurements in addition to birth weight, as is described in more detail in Section 2.3.7. The first EFW is used to establish the upper and lower bounds for the initial condition $(\tilde{t}_0, \tilde{W}_{f_0})$ used for solving equation (2.21). In this study, on average, the first

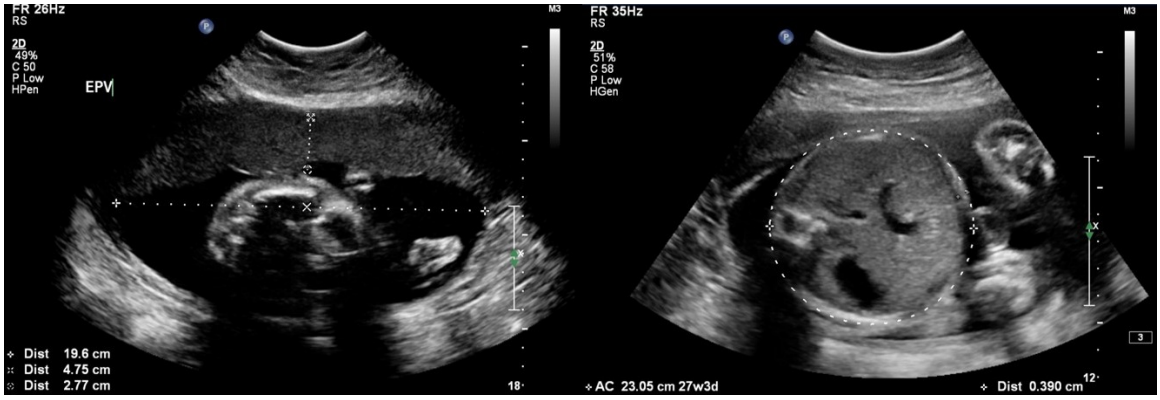

Figure 2.2: Example of an Ultrasound Report for Establishing Estimated Measurements of Estimated Fetal Weight (EFW), Estimated Placental Volume (EPV), and Fetal Body Composition of a Representative HMZ Participant.

ultrasound measurement was taken at 14 weeks gestation, followed by five additional measurements each every four weeks through 34 weeks gestation. Infant's birth weight was measured immediately after delivery.

2.3.2 Estimated Placental Volume

As the case with EFW, up to six ultrasound measurements are used to obtain the estimated placental volume (EPV) measurements using the Azpurua *et al.*, 2010 approximation method, for which a number of simplifying assumptions have been made. As detailed in Section 2.3.8, EPV measurements are incorporated in the estimation cost function with lower emphasis than EFW measurements. This is justified given the following:

1) Absence of EPV measurements at or near birth. The bias that can result from equally emphasizing EPV measurements with EFW in (2.32) given the missing value at birth is crucial since fitting to earlier measurements only will tend to produce an exponential growth profile that can be, misleadingly, well-captured with the modified logistic equation in (2.29).

2) The Azpurua *et al.*, 2010 EPV estimation method using 2D ultrasound mea-

surements (similar to Figure 2.2) provides a rather simplified approximation (assumes spherical topology) that is mainly targeted for establishing EPV in the first and second trimesters of pregnancy for patients with normal BMI index (in this study, participants are either overweight or obese); this approximation can become exceedingly inaccurate at advanced gestational ages due to remaining technical difficulties associated with existing ultrasound technology.

3) The EPV approximation method in Azpurua *et al.* does not estimate standard errors; only 10^{th}, 50^{th}, and 90^{th} percentile trajectories are given.

2.3.3 Fetal Body Composition

Studying ultrasound reports similar to Figure 2.2 (namely, anterior abdominal wall thickness and abdominal circumference) also produced at least two acceptable estimates per participant for the fetal % body fat using the correlation presented by Bernstein and Catalano, 1991. While the number of estimates can be as many as available ultrasounds, it is known that this estimate becomes more reliable at advanced gestational ages, and therefore we only consider a subset of the ultrasound measurements for the estimation of fetal % body fat.

2.3.4 Glycemic Index

Glycemic index (GI) was estimated using food items and portion sizes reported in a smart phone application. For each food, carbohydrate content (g) of the reported portion size was determined using the USDA Food and Nutrient Database for Dietary Studies (FNDDS) 2013-2014 data set. Next, a GI value for each food was determined by matching foods to the database generated by Flood *et al.*, 2006. For foods without an exact match, the GI value of the closest matching item was used. Estimated GI

of each day was then calculated as the average GI of all foods consumed in a day, weighted by their contribution to total carbohydrate intake for that day, i.e.

$$g(t) = \left. \frac{\sum_{i=1}^{n} \text{food } i \text{ GI} \times \text{food } i \text{ carbohydrate [g]}}{\sum_{i=1}^{n} \text{food } i \text{ carbohydrate [g]}} \right|_t \tag{2.30}$$

In equation (2.21), the glycemic index, $g(t)$, is understood as a key variable for estimating fetal energy intake. One can generally presume a time-varying profile of $g(t)$ on a daily scale; however, given that the collected $g(t)$ time series shows to be stationary with a low variance, one can assume a constant $\overline{g}$ value drawn from the average of all available average daily estimates, $\overline{g}$. Table 2.2 lists the fractional average and standard deviation of estimated daily glycemic index values for the four participants presented in this study.

Table 2.2: Mean and Standard Deviation Values of Daily Glycemic Index Estimations for Four Representative HMZ Participants

Participant	$\overline{g}$	σ_g
A	0.5478	0.0461
B	0.5441	0.0653
C	0.5806	0.0644
D	0.5763	0.0835

2.3.5 Maternal Physical Activity

As noted in section 2.2.3, equation (2.22) characterizes the assumed simple, linear dependence of the fetal energy balance model in (2.21) on maternal physical activity. It is assumed that physical activity moderates the energy intake to the fetus by regulating the placental function (e.g., through blood flow; Ferraro *et al.*, 2012). In the HMZ study, intensive objective assessment of physical activity is carried out using wrist-worn activity tracker. Missing and implausible physical activity measurements are imputed with mean replacement. These data are also used to establish the estimated daily maternal energy intake in equation (2.31).

2.3.6 Maternal Energy Intake

The daily maternal energy intake variable, $m(t)$, can be reliably estimated with the availability of daily maternal weights and estimated energy expenditure data; the latter are estimated by correlating with daily physical activity and estimated/measured resting metabolic rates. As presented in Guo *et al.*, 2016, back-calculated maternal energy intake from measured daily maternal weights and physical activity measures is considered:

$$m(t) = \frac{-W_m(t+2) + 8W_m(t+1) - 8W_m(t-1) + W_m(t-2)}{12TK_1} - \frac{K_2}{K_1}\left[PA(t) + RMR(t)\right] \tag{2.31}$$

where K_1 and K_2 are gains (coefficients) that map changes of daily energy intake and physical activity, respectively, into maternal weight gain/loss; T is the sampling time; $PA(t)$ and $RMR(t)$ are the maternal daily physical activity and resting metabolic rate, respectively. To reduce the significant variability in equation (2.31), it is necessary to smooth the weight measurement $W_m(t)$. A 9-day moving average filter is considered for all participants, except for participant D where a 13-day moving average filter is considered.

2.3.7 Model Estimation Problem Formulation

In this section, we establish a problem formulation for the least squares objective from which, with the presence of sufficient estimation and validation data, model parameters can be estimated and validated using nonlinear regression. Next, we describe in more detail how emphasis is split between different measured variables, and how the nonlinear optimization solver is initialized. Finally, in the results section, we present simulations of the estimated individual models and list the mean value and standard deviation associated with all model parameters.

The parameter estimation problem statement is formulated as a constrained optimization problem. The prediction error is minimized over estimation data using a non-linear least squares objective. For model estimation, using a total of N EFW measurements inferred from ultrasound reports (similar to the example sonographic images shown in Figure 2.2, including birth weight), M EPV measurements, and L estimated body composition data points, the approach considered is to solve

$$\begin{aligned} \min_{\theta} \quad & \varepsilon^T Q \varepsilon \\ \text{s.t.} \quad & \theta_{lb} < \theta < \theta^{ub} \end{aligned} \tag{2.32}$$

where

$$\varepsilon = \begin{bmatrix} \Delta W_f(1) & \cdots & \Delta W_f(N) & \Delta P(1) & \cdots & \Delta P(M) & \Delta FM_f(1) & \cdots & \Delta FM_f(L) \end{bmatrix}^T$$

$$Q = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_{N+M+L} \end{bmatrix}$$

$$\theta = \begin{bmatrix} \alpha & e_{FM_f} & e_{FFM_f} & a_P & b_P & c_P & a_{f_r} & b_{f_r} & c_{f_r} & \tilde{W}_{f_0} \end{bmatrix}^T$$

$\Delta W_f(t) = EFW(t) - W_f(t)$, $\Delta P(t) = EPV(t) - P(t)$, and $\Delta FM_f(t) = \hat{FM}_f(t) - FM_f(t)$ with $EFW(t)$, $EPV(t)$, and $\hat{FM}_f(t)$ denoting the estimated ultrasound measurements of the fetal weight, placental volume, and fetal fat mass at day t, respectively; Q is a positive semi-definite weighting matrix used to establish the desired emphasis for model estimation. $W_f(t)$ is obtained from the numerical solution of the following fetal model

$$\tau_f(W_f)\frac{dW_f(t)}{dt} + W_f(t) = K_f(t)m(t)P(t), \quad W_f(\tilde{t}_0) = \tilde{W}_{f_0} \tag{2.33a}$$

$$\tau_f(W_f) = \frac{\lambda_{FM_f} f_r(W_f)/e_{FM_f} + \lambda_{FFM_f}\left[1 - f_r(W_f)\right]/e_{FFM_f}}{\mu} \tag{2.33b}$$

$$K_f(t) = \frac{\gamma(t)\overline{g}}{\mu}, \quad \gamma(t) = \alpha PA(t) + \beta, \quad \alpha \leq 0, \quad \beta \geq \gamma > 0 \qquad \forall\ t \geq 0 \tag{2.33c}$$

$$P(t) = c_P\left[\frac{1}{1 + e^{-a_P(t-b_P)}} - \frac{1}{1 + e^{a_P b_P}}\right] \tag{2.33d}$$

$$f_r(W_f) = a_{f_r} c_{f_r} \frac{e^{-a_{f_r}\left(W_f(t) - b_{f_r}\right)}}{\left[1 + e^{-a_{f_r}\left(W_f(t) - b_{f_r}\right)}\right]^2} \qquad \forall\ t \geq 0 \tag{2.33e}$$

with $\tilde{t}_0 > 0$ (initial time of simulation) and $m(t)$ per equation (2.31). Lower and upper parameter bounds, θ_{lb} and θ^{ub}, are known *a priori*. The physical activity parameter α is constrained as shown in equation (2.33c). In this work, for purposes of simplicity, the value of the β parameter is fixed at 0.000234 ml^{-1}, which is equal to the estimated nominal value of γ in Thomas *et al.*, 2008. Thermodynamic efficiencies e_{FM_f} and e_{FFM_f}, by definition, range from 0 to 1. Also, given the *strict* growth of both $P(t)$ and $FM_f(t)$ profiles, the parameters $a_P, b_P, c_P, a_{f_r}, b_{f_r}$, and c_{f_r} are bounded below at 0, and are unbounded above.

The optimization is initialized using nominal parameter values/ranges drawn from literature. For example, Christiansen *et al.* Christiansen *et al.*, 2005; Noblet *et al.*, 1999 reports values for thermodynamic efficiencies drawn from animal studies; Thomas *et al.*, 2008 gives an estimate for the conversion parameter, $\gamma(t)$; Demerath *et al.*, 2016 provides fat and fat-free mass profiles from preterm infants that are used for initializing a_{f_r}, b_{f_r} and c_{f_r} using standard regression; finally, also by similar means, EPV measurements calculated from our ultrasound data are used for initializing a_P, b_P, and c_P. In the following section, we report in additional detail on the final set of parameter values used for solver initialization.

2.3.8 Relative Weights & Initialization

In this section, the specific relative weights (λ_i in the diagonal Q matrix in equation (2.32)) are presented for each participant. In addition, the specific initialization points (initial guesses) are also established in this section. It must be noted that given the limited amount of estimation data and the non-convexity of the optimization problem, the non-linear least squares solver becomes increasingly sensitive to relative weights and proper initialization as multiple local minima are expected. To avoid undesired solutions, solver features such as *multistart* can be used (Ugray

et al., 2007).

Judicious selection of λ_i values is important for establishing an effective estimation cost function for each of the HMZ participants evaluated with this method. In the selection of λ_i values, output emphasis, scaling, number of measurements, and measurement standard errors are all taken into consideration. While each data point can have its specific assigned λ_i weight, we group measurements per model state (i.e., EFW, EPV, fetal FM) with one relative weight as $\lambda_{EFW} : \lambda_{EPV} : \lambda_{FM_f}$. For participant A, the established ratios are $1 : 0.5 : 1$, whereas for participants B, C, and D the ratios are $1 : 0.3 : 1$.

Table 2.3 lists established initialization points for the studied HMZ participants. In the selection of these initializations, approximations from the literature, actual measurement values, and multiple iterations are all influencing factors. More specifically, initial guesses for e_{FM_f} and e_{FFM_f} were drawn from Christiansen *et al.*, 2005 followed by multiple iterations (multiple solutions); a_P was drawn from Thomas *et al.*, 2008; b_P, c_P, a_{f_r}, b_{f_r}, and c_{f_r} were initialized from examining the actual measurements followed by multiple iterations; finally, the initialization of α was established after multiple iterations.

Table 2.3: Initialization Points for Four Representative HMZ Participants

	$\alpha \times 10^8$	e_{FM_f}	e_{FFM_f}	a_P	b_P	c_P	a_{f_r}	b_{f_r}	c_{f_r}
A	-0.5	0.77	0.11	0.03	175	1281	0.47	9.56	9.47
B	-0.5	0.77	0.11	0.03	175	864	0.47	9.56	12.30
C	-0.5	0.44	0.15	0.03	175	774	0.47	9.56	8.99
D	-0.5	0.77	0.11	0.03	175	760	0.47	9.56	10.98

2.3.9 Estimation Results

In this section, for each of the examined HMZ participants, qualitative and quantitative model fit to data are presented from simulations when actual measured inputs

are applied to the model. In addition to the intrauterine fetal weight (primary model state), other model states (i.e., placental volume, body composition) and the evolution of intermediate constructs over time (e.g., $e_f(t)$, $\tau_f(t)$, and $K_f(t)$) are also shown. Finally, estimated model parameters tabulated in Table 2.4 are discussed.

Figures 2.3-2.7 feature simulations of the estimated models for one intervention participant (participant A) and one control participant (participant B). Overall, the goodness of fit does not appear to differ across intervention and control participants. In Figures 2.3-2.7, the simulation start time is selected to match the day of the first ultrasound measurement; the simulation is carried out through the reported actual day of birth. In these simulations, measurements of the two model inputs, i.e., maternal energy intake (back-calculated EI) and maternal PA (direct measurements), are displayed. In addition, the model states, i.e., fetal weight, placental volume, and body composition, are plotted and contrasted against estimated ultrasound measurements to qualitatively demonstrate the goodness of fit. Moreover, in Figure 2.9 and Figure 2.11 featuring the time-varying profiles of τ_f, e_f, and K_f; it can be seen that, across all individuals, both τ_f and e_f appear to exponentially increase over time as the fetus continues to grow. It is noted that Thomas *et al.*, 2008 provides a significantly higher estimate for the overall efficiency ($e_f = 0.799$) than the estimated ranges from our data (approximately, in the 0.1-0.4 range). Finally, Table 2.4 summarizes the estimated model parameters with mean and standard deviation values for the examined participants.

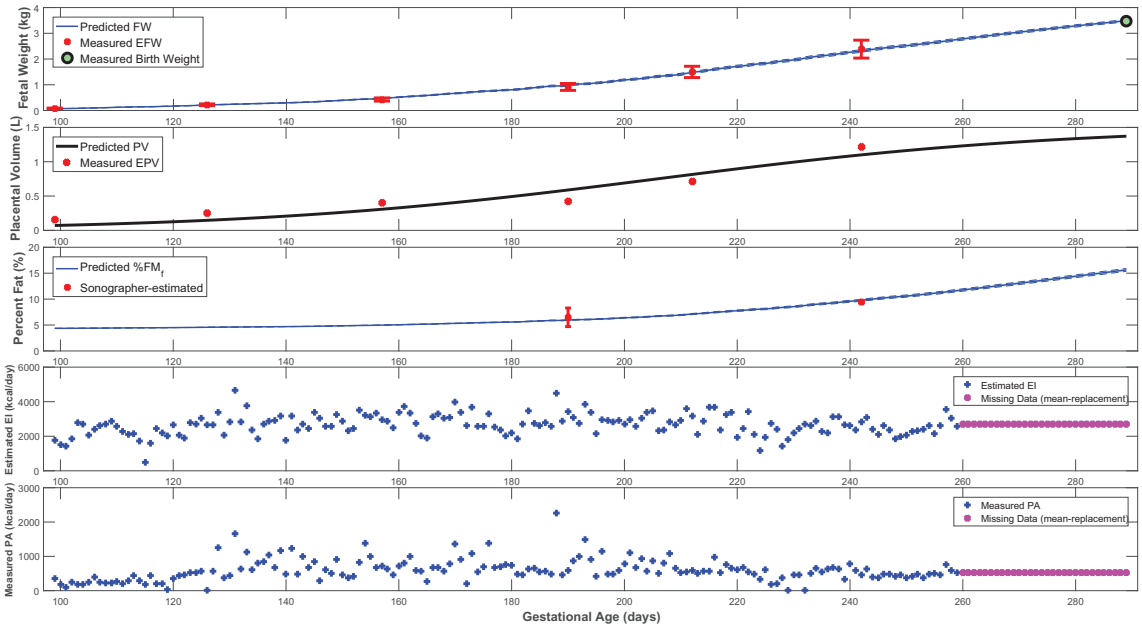

Figure 2.3: Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Intervention Participant (Participant A). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth.

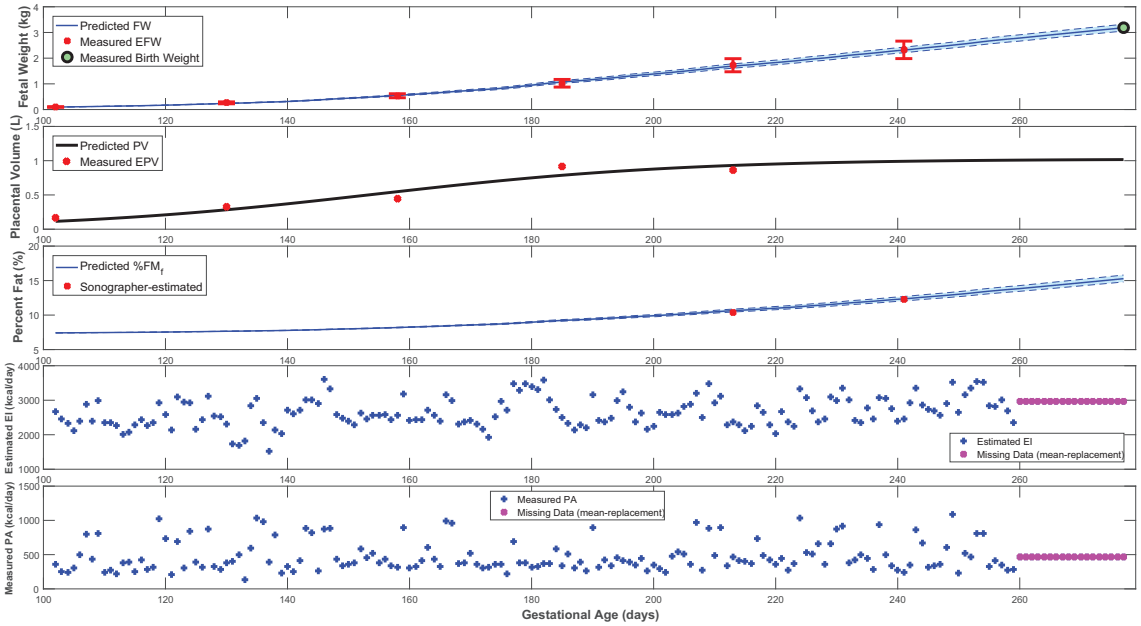

Figure 2.4: Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Control Participant (Participant B). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth.

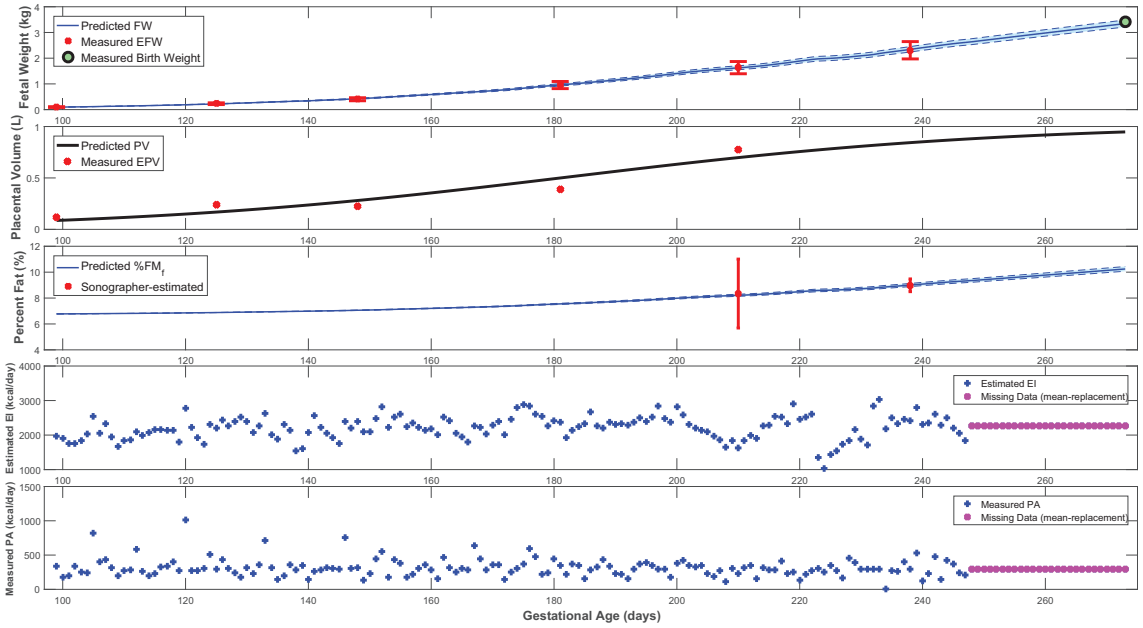

Figure 2.5: Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Intervention Participant (Participant C). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth.

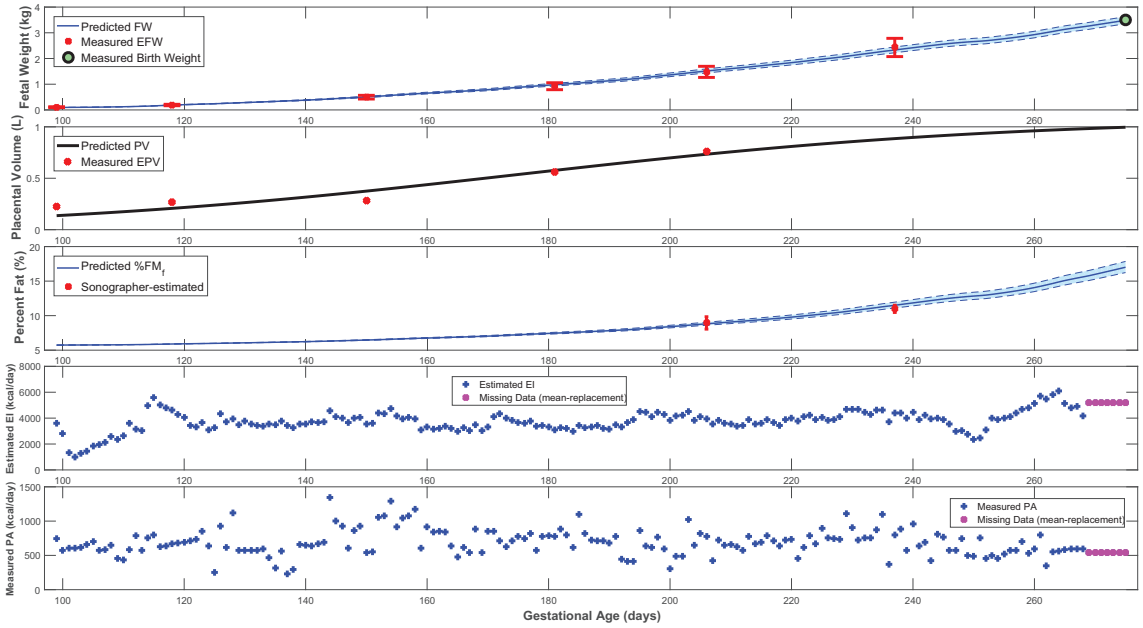

Figure 2.6: Time-domain Response (Fetal Weight, Placental Volume, and Fetal % Body Fat) with Energy Intake and Physical Activity for a Representative HMZ Control Participant (Participant D). Simulation Starts at the Day of First Ultrasound Measurement and Ends at Birth.

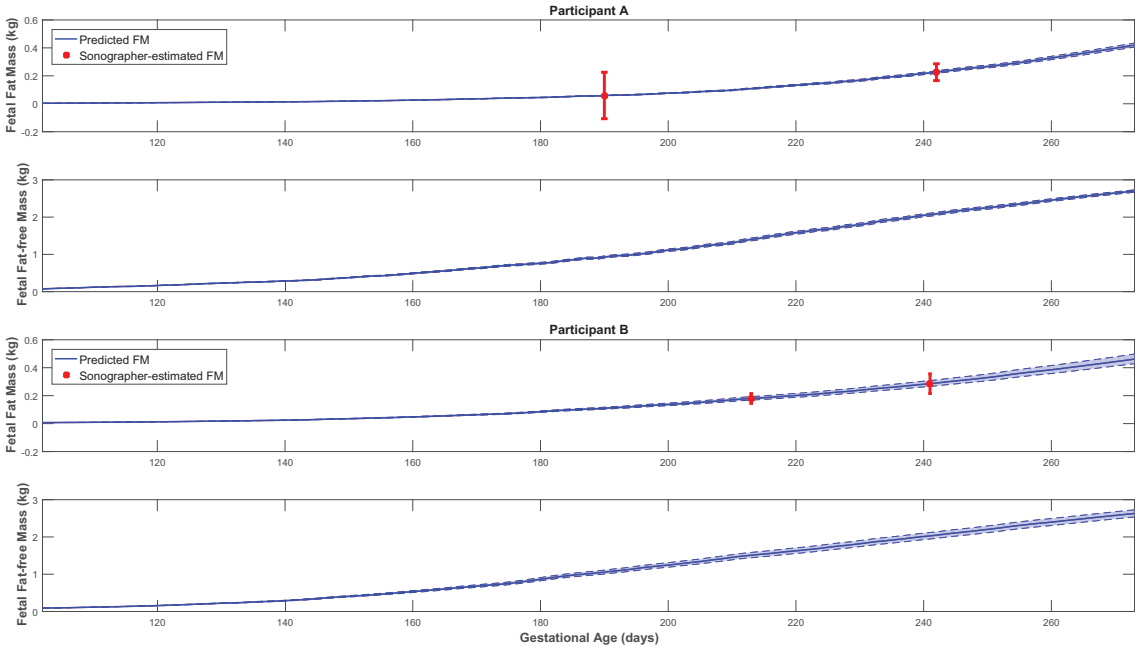

Figure 2.7: Fetal Fat Mass and Fat-free Mass Growth Profiles over Time for Representative HMZ Participants (Participants A and B).

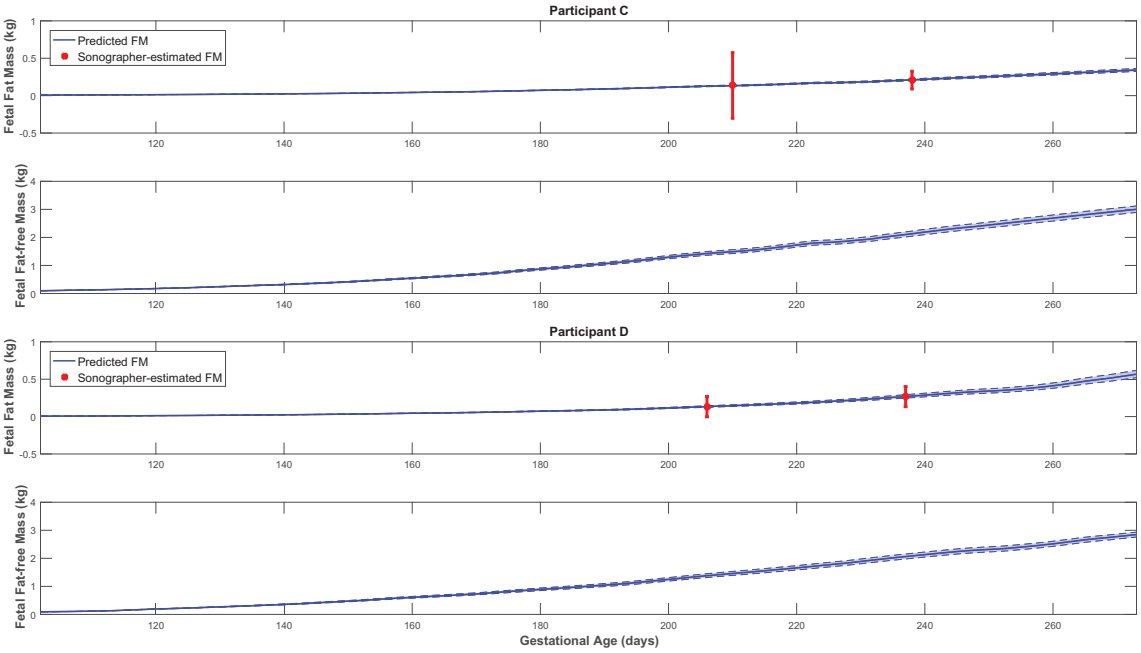

Figure 2.8: Fetal Fat Mass and Fat-free Mass Growth Profiles over Time for Representative HMZ Participants (Participants C and D).

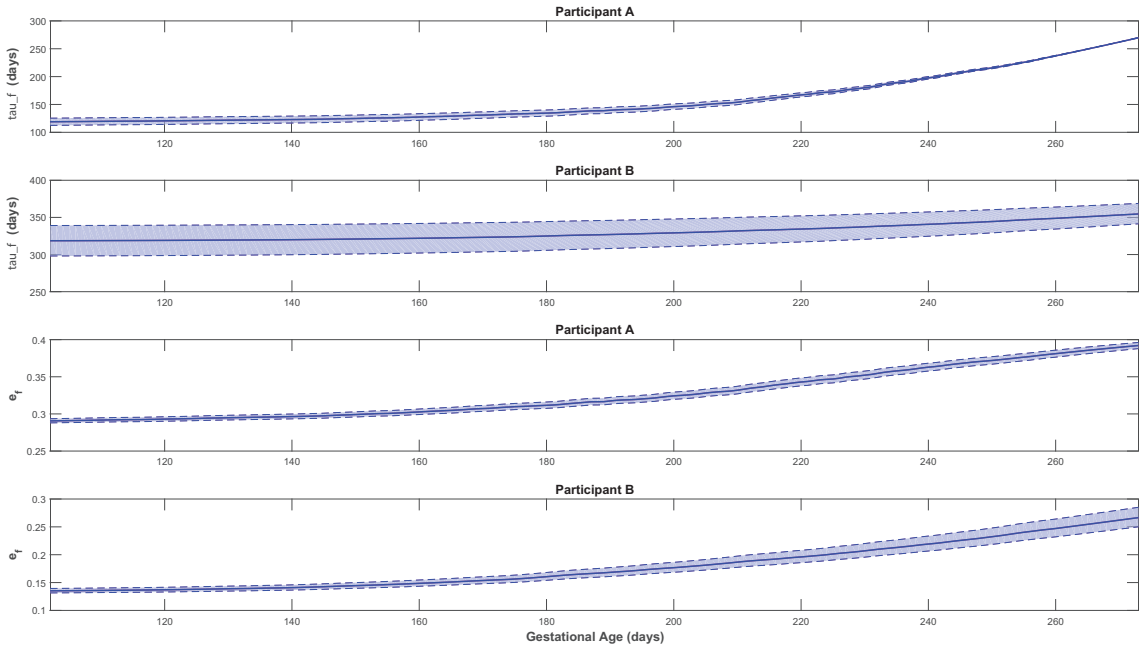

Figure 2.9: Time-varying τ_f and e_f for Representative HMZ Participants (Participants A and B).

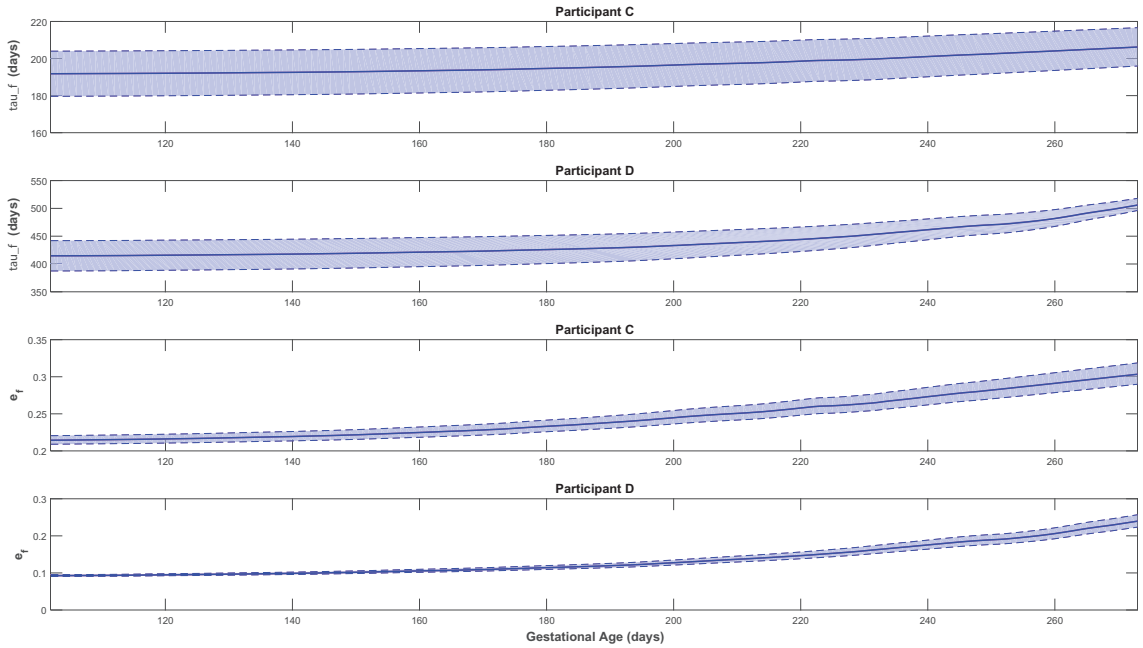

Figure 2.10: Time-varying τ_f and e_f for Representative HMZ Participants (Participants C and D).

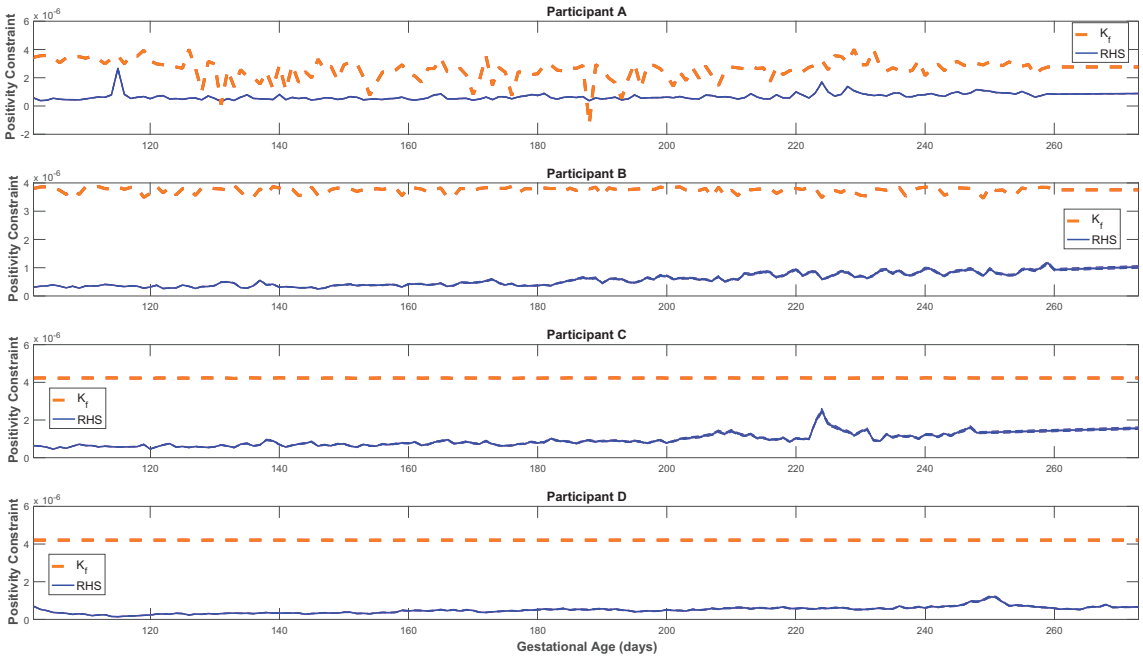

Figure 2.11: Time-varying Gain and the Establishment of Positive Fetal Energy Balance for Representative HMZ Participants (See equation (2.23)).

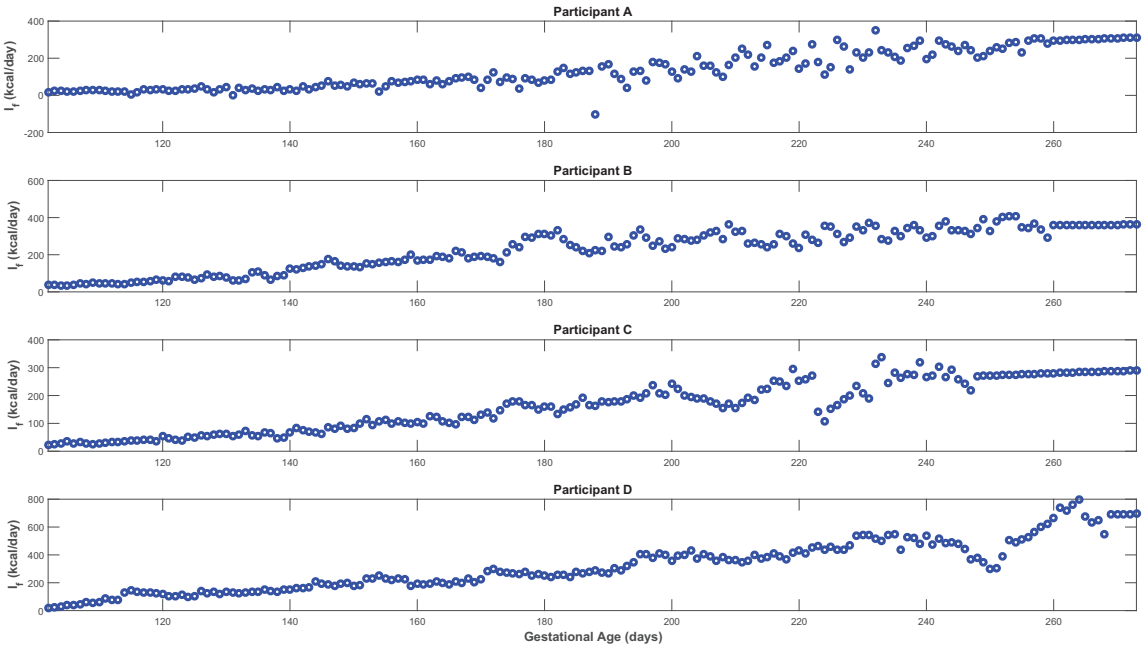

Figure 2.12: Predicted Time-domain Profile of Fetal Energy Intake $I_f(t)$ for Representative Hmz Participants (See equation (2.15)).

Table 2.4: Estimated Model Parameter Values for Four Representative Hmz Participants with Mean and Standard Deviation (SD)

	$\alpha \times 10^8$	e_{FM_f}	e_{FFM_f}	a_P	b_P	c_P	a_{f_r}	b_{f_r}	c_{f_r}
A	-13.70	0.44	0.24	0.027	206.2	1519.1	0.655	7.86	11.33
B	-2.82	0.60	0.07	0.039	154.3	1026.8	0.441	9.71	12.30
C	-0.31	0.82	0.12	0.028	182.3	1031.0	0.267	12.10	6.84
D	-0.06	0.42	0.06	0.024	174.8	1101.7	0.577	8.22	11.27
Mean	-4.22	0.57	0.12	0.030	179.4	1169.7	0.485	9.47	10.44
SD	6.44	0.18	0.08	0.007	21.4	235.5	0.170	1.93	2.44

2.3.10 Model Validation

The work of Thomas *et al.*, 2008 established their model performance using compiled data from various sources in the literature as summarized in Table 2.5; plots evaluating their placental volume model performance against three data points for both exercising and non-exercising groups were also provided in the paper. However, with longitudinal data from *Healthy Mom Zone*, the performance and validation of the improved model in this work are determined by goodness-of-fit metrics as well as contrasting diverse estimated model features such as structure, parameter ranges, and output profiles against prior knowledge from existing literature.

Table 2.5: Performance Summary of the Model Developed in Thomas *et al.*, 2008

	Prediction (Birth Weight)	Prediction (Fat Mass)
Low Glycemic Diet	96.40%	86.84%
Runner Group	98.55%	97.37%
Non-exercising Group	100%	73.81%

First, all simulated fetal weight and placental volume growth profiles in Figures 2.3, 2.4, 2.5, and 2.6 are plausible and consistent with expected growth profiles from literature (Hadlock *et al.*, 1984, 1985, 1991; Arleo *et al.*, 2014). Comparing model predictions against the experimentally observed data (ultrasounds), a summary of individualized model outputs fit against available data is presented in Table 2.6. The Normalized Root-Mean-Square Error (NRMSE) is defined as

Table 2.6: Summary of the Goodness-of-fit from Various Metrics for Four Representative HMZ Participants

	$\text{NRMSE}_{\text{EFW}}$	R^2_{EFW}	R^2_{EPV}	$R^2_{FM_f}$
Participant A	0.9619	0.9986	0.8953	1.0000
Participant B	0.9763	0.9994	0.9008	0.9925
Participant C	0.9608	0.9985	0.8969	0.9997
Participant D	0.9763	0.9994	0.9174	1.0000

$$\text{NRMSE}_{\text{EFW}} = 1 - \frac{\|EFW(t) - W_f(t)\|_2}{\|EFW(t) - \overline{EFW}\|_2} \tag{2.34}$$

is considered as the primary metric for establishing the model goodness of fit against the HMZ data. $W_f(t)$ is the simulated output, $EFW(t)$ is the measured output, $\overline{EFW}$ is the mean of all measured $EFW(t)$ values, and $\|\cdot\|_2$ denotes the l_2-norm. R^2_{EFW}, R^2_{EPV}, and $R^2_{FM_f}$ denote the coefficients of determination for fitting to estimated fetal weights, placental volumes, and fetal fat mass *in utero*, respectively. For a qualitative evaluation of model fit, the reader may refer to Figures 2.3-2.7.

Further giving validity to our model is that the estimated e_{FM_f} values are consistently larger than e_{FFM_f}, in agreement with reported patterns only available from animal studies (Christiansen *et al.*, 2005). Moreover, the mean estimated value of the placental volume growth rate parameter a_P (with a narrow standard deviation of 0.003) matches the reported and validated value in Thomas *et al.*, 2008; Orzechowski *et al.*, 2014: $r = 0.03$. Furthermore, from Figures 2.3, 2.4, 2.5, and 2.6, predicted % body fat at birth approximately ranges from 10 to 18%, which fall into the typical ranges reported in literature (Widdowson and Spray, 1951; Schwartz and Galan, 2003; Demerath and Fields, 2014; Bernstein and Catalano, 1991). In agreement with Demerath *et al.* (2016) Demerath *et al.*, 2016, Figures 2.7 and 2.8 show that predicted $FFM_f(t)$ profiles can be described as linear, while the $FM_f(t)$ are curvilinear (linear-exponential).

Finally, Figure 2.11 confirms that, except for only two brief instances in Participant A's simulations, all estimated models satisfy the constraint in equation (2.23)

and hence validates the positive energy balance assumption throughout gestation. From Figure 2.12, one can observe the estimated rate of fetal energy intake $I_f(t)$ (note the negative $I_f(t)$ values in the two instances where Participant A's positivity constraint is violated); comparing this to the maternal energy intake $m(t)$ ('Estimated EI' in Figures 2.3, 2.4, 2.5, and 2.6) provides support for the assumption of a well-nourished mother.

2.4 Chapter Summary

In conclusion, a dynamical systems model of intrauterine growth has been developed from first-principles, relying on the first and second laws of thermodynamics. This proposed model provides a rigorous yet more simple formulation than the fetal energy balance model of current literature. In the parameter estimation of this model, a non-linear least squares, constrained multi-objective optimization problem was formulated and guided by *a priori* knowledge of ranges of model parameters. For the first time (to the authors' knowledge), estimates (and an estimation method) for the thermodynamic efficiencies governing the formation of new tissues of human fetuses are established. This developed model has been estimated and validated against ultrasound measurements provided from the *Healthy Mom Zone* study; despite the explained challenges with the estimation measurements, predictions follow from this model show good agreement with the data.

The availability of more intensive estimation and validation datasets (i.e., datasets with more frequent measurements) in a future study should create opportunities for parameter refinement and increased model understanding. More intensive measurements will allow for further investigation of the contribution of maternal body components in fetal nutrition (see equation (2.14)). Moreover, additional ultrasound measurements (particularly closer to delivery) may allow estimation of a less biased fetal

model described in equation (2.24). A better theoretical understanding of mechanisms behind the evolution of placental volume and the rate of fetal fat mass deposition is needed. This work considered a linear dependence of placental function on maternal physical activity; in future work, a more developed characterization of the influence of maternal physical activity may generate more resilient models: models with good predictions when input levels are far from those used in model estimation. Furthermore, a broader future goal is to use a combination of more experimental data and increased physiological understanding to reduce the modeling assumptions (outlined in Section 2.2) as much as possible.

Finally, the aim of this work (achieved using a limited number of HMZ intervention and control participants) was to develop a more comprehensive energy balance model for fetal weight gain derived from first-principles modeling that can be validated through data. These aims were facilitated by the availability of intensive, longitudinal participant data from the HMZ intervention. Model estimation and validation efforts for the additional HMZ participants ($N = 32$) could enable making conclusions regarding participant differences and intervention versus control outcomes, which was not the scope of this work. However, studying group differences (intervention vs. control) is a subject of current and future research.

Chapter 3

BLACK-BOX IDENTIFICATION OF *JUST WALK*: TOWARD ESTIMATING DYNAMICAL SYSTEMS MODELS OF SOCIAL COGNITIVE THEORY

3.1 Background

One recent emerging application of system identification and control theory is the design of optimized interventions in health behavior. In designing adaptive interventions (Rivera *et al.*, 2007), a key consideration is the ability to estimate *personalized* behavior models that identify both individual-invariant and individual-variant dynamics. Parsimonious modeling, guided by *a priori* knowledge, is therefore crucial to accomplishing this. Using control systems engineering principles, behavior change theory can be utilized to develop models and decision frameworks for interventions that promote physical activity (PA) among sedentary individuals. One example is Social Cognitive Theory (SCT), proposed by Bandura (Bandura, 1986), which is among the leading theories of behavior change. The work of Martín *et al.* (Martín *et al.*, 2014) established a dynamical systems fluid analogy model that captures key SCT concepts. Figure 1.1 represents a simplified fluid analogy dynamical system model of Social Cognitive Theory. SCT includes an extended list of potential constructs associated with predicting complex behavior dynamics, and whose effect needs to be accounted for during modeling. Consider, for example, the *Environmental Context* construct in Figure 1.1; this can include weather, busyness, stress, weekday, mood, and several other known and unknown variables.

Following an experimental design methodology based on system identification principles (Martín *et al.*, 2015), a unique single-subject intervention study, *Just Walk*,

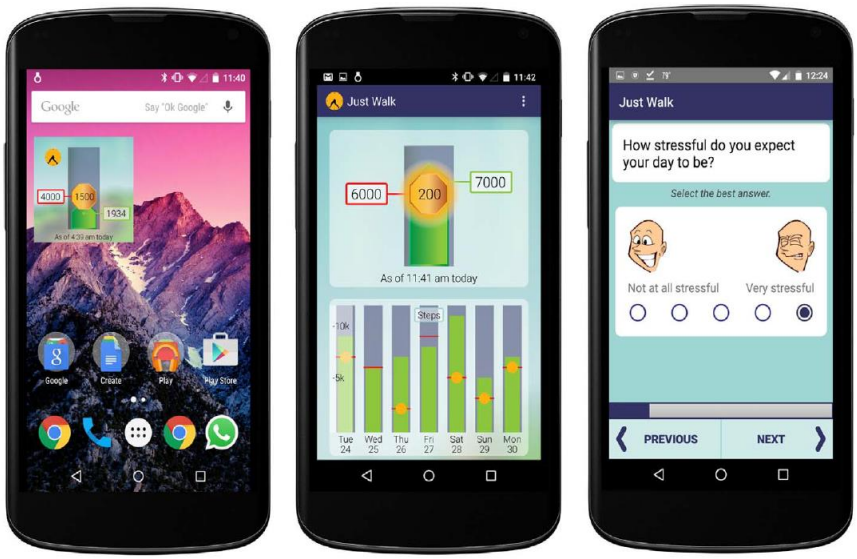

Figure 3.1: Screenshots of the *Just Walk* Mobile Application (Phatak *et al.*, 2018).

was performed. Section 3.2 introduces the *Just Walk* intervention and experiment execution, followed by a brief description on the input signal design approach using orthogonal multisine excitations in Section 3.3.

Section 3.4 discusses the use of an unconventional black-box approach that provides key insights into the dynamics of the intervention participants, and will ultimately be instrumental in accomplishing semi-physical identification of SCT models (Figure 1.1). Section 3.5 outlines a number of important conclusions and future directions on input signal design, modeling, identification, and intervention design.

3.2 Description of the *Just Walk* Intervention

Just Walk was developed as an adaptive walking intervention app for sedentary, overweight adults. It was designed primarily as a tool to generate individualized computational models for understanding PA behavior via system identification. The intervention system included a front-end Android app, *Just Walk* (Figure 3.1), a backend server, and an activity tracker (Fitbit Zip) to objectively measure PA and

automatically sync with the smartphone application. Participants were recruited nationally to partake in a walking intervention and receive daily step goals via the *Just Walk* app, and daily announced points were granted if the goals were achieved that day; granted points were converted into Amazon gift cards after a certain threshold was reached. Participants were also required to complete a series of daily morning and evening ecological momentary assessment (EMA; Shiffman *et al.*, 2008) measures (e.g. confidence in achieving goal, predicted business for that day, previous night's sleep quality, etc.) for the entire duration of the study.

The study duration was 14 weeks, including an initial two-week baseline period in which no step goals were delivered. Each participant's step goals were then based on their median daily step value as calculated from the 14-day baseline period. The step goals were designed to establish a mechanism for individualizing the definition of an "ambitious, but doable" step range. All PA data were collected from the Fitbit Zip (provided to participants as a part of the study) and stored both locally and in Fitabase (Small Steps Labs, San Diego, CA, USA). Participants were generally healthy, inactive, 40-65 years old, with a body mass index (BMI) of 25-45 kg/m^2, who currently owned an Android phone capable of connecting to a Fitbit Zip via Bluetooth 4.0, and were willing to engage with the mHealth intervention for 14 weeks.

3.3 Input Signal Design of *Just Walk*

The input signal design procedure utilized in the *Just Walk* study was designed using deterministic yet "pseudo-random" signals that are orthogonal in the frequency domain. The procedure is described in detail in Martín *et al.*, 2015. In *Just Walk*, *Goals* establish the desired behavior in a quantitative form, while *Expected Points* are the daily available points announced each morning that are granted upon goal achievement. *Goals* and *Expected Points* are two manipulated input signals u_n gen-

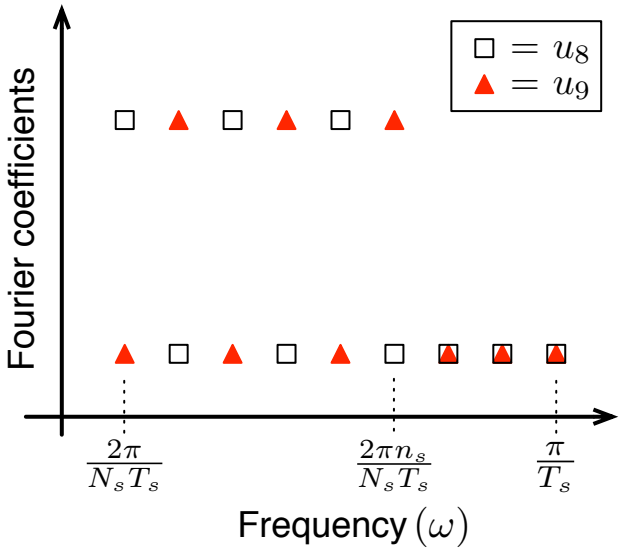

Figure 3.2: Conceptual Representation of a "Zippered" Spectra Design for $n_u = 2$ Design Inputs, and $n_s = 6$ Harmonic Frequencies (Freigoun *et al.*, 2017).

erated from a multisine signal,

$$u_n(k) = \lambda_n \sum_{j=1}^{N_s/2} \sqrt{2\alpha_{[n,j]}} \cos(\omega_j k T_s + \phi_{[n,j]})$$
$$\omega_j = \tfrac{2\pi j}{N_s T_s}, \quad k = 1, \ldots, N_s \tag{3.1}$$

where λ_n is the scaling factor, N_s is the number of samples per period, T_s is the sampling time. For the j^{th} harmonic of the signal each variable has the following meaning: $\alpha_{[n,j]}$ is a factor used to specify the relative power of the harmonic, ω_j is the frequency, and $\phi_{[n,j]}$ is the phase. To obtain independent transfer function and uncertainty estimates, factors $\alpha_{[n,j]}$ are chosen to excite input signals orthogonally in frequency. Two signals are orthogonal if a nonzero Fourier coefficient at a specific frequency in one signal implies a zero-valued Fourier coefficient at the same frequency for the other; this is called a "zippered" spectra design, an idea introduced in Rivera *et al.*, 2009. A conceptual representation of the "zippered" design is presented in Figure 3.2. For n_u design inputs and n_s independently excited sinusoids the Fourier

coefficients are specified as

$$\alpha_{[n,j]} = \begin{cases} 1 & \text{if } j = n_u(i-1) + (n-7) \\ & \text{for } i = 1, 2, \ldots, n_s \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

Using the ω_j frequencies defined in (3.1) and the Nyquist–Shannon sampling theorem, the following bound for N_s is defined:

$$N_s \geq 2n_s \tag{3.3}$$

If $n_s = 6$ excited sinusoids are selected for the $n_u = 2$ design inputs, then from applying (3.3), $N_s = 16$ days (selected) is a feasible option. Phases $\phi_{[n,j]}$ are selected to minimize the crest factor of the signal using the approach proposed by Guillaume *et al.*, 1991.

In applying this design methodology for *Just Walk*, amplitudes for input signals (u_8 and u_9 in Figure 3.3) were chosen relying on experiences from previous studies (King *et al.*, 2013; Adams *et al.*, 2013) designed to obtain an expected profile of PA. The maximum number of step goals was selected as a factor of the initial baseline level of PA. For most cases in this experimental design, this factor was equal to 2; however, it was varied if the actual baseline step level of individuals was too high or low. Specifically, if participant's baseline median steps were below 3,000, then the range for the goals was between 1 and 2.5 of their baseline median steps, to increase the likelihood of "ambitious" goals. If baseline median steps were greater than 7,500 steps, then the range was set between 1 and 1.75 (to reduce the likelihood of overly ambitious goals, such as 15,000 steps in one day). In addition to the two manipulated input channels, a large set of disturbances were also measured using mHealth technologies.

Overall experimental duration beyond the baseline varied between five to six cycles for each participant. A time series plot for a representative participant that depicts the behavior and seven inputs is shown in Figure 3.3.

3.4 ARX Model Estimation & Validation

In this section, black-box modeling strategies used for *Just Walk* are outlined, and results from fitting Auto Regressive with eXogenous input (ARX) parametric models (Ljung, 1999) are presented. As noted, identifying optimal ARX models (decisions on model inputs and model order) will play a pivotal role in ultimately identifying *personalized* semi-physical (grey-box) models that are informed by well-established behavior theories (Martín *et al.*, 2014). Prior to ARX estimation, standard nonparametric modeling tasks such as correlation analysis have been informative. Because the *Just Walk* study included a wide array of input/output measurements, results from input-output and input-input correlation analyses have been useful (Phatak *et al.*, 2016; Hekler, 2015; Korinek *et al.*, 2018); for brevity, these are not included in this work. Incorporating all measured disturbances for estimating an SCT behavior model (particularly the *Environmental Context* construct in Figure 1.1) can be computationally demanding, may pose *identifiability* challenges, and will require large informative datasets that are typically difficult to gather from a practical standpoint in research involving human subjects.

Preprocessed data are fitted to an ARX model structure ARX-[n_a,n_{b_1},...,$n_{b_{n_u}}$, n_{k_1},...,$n_{k_{n_u}}$], which can be expressed in the following concise form:

$$y(k) + \sum_{l=1}^{n_a} a_l y(k-l) = \sum_{j=1}^{n_u} \sum_{i=1}^{n_{b_j}} b_{ij} u_j \left(k - n_{k_j} - i\right) + e(k) \tag{3.4}$$

where $y(k)$ is the measured output (e.g., steps/day), $u_j(k)$ is the measured input j, $e(k)$ is the prediction error, all measured/estimated at day k. The ARX model

in (3.4) is estimated by using regression. ARX parameter estimation constitutes a linear least-squares regression problem (Ljung, 1999) and has attractive statistical properties such as consistency. Figure 3.3 illustrates an example contrasting the difference between actual output measurements and the prediction from a 7-input ARX model with the structure in (3.4); a detailed discussion of black-box modeling strategies used in this work follows. To quantify model fits, the normalized root mean square error (NRMSE) fit index is used

$$\text{model fit } (\%) = 100 \times \left(1 - \frac{\|y(k) - \hat{y}(k)\|_2}{\|y(k) - \bar{y}\|_2}\right) \tag{3.5}$$

$y(k)$ is the measured output, $\hat{y}(k)$ is the simulated output, $\bar{y}$ is the mean of all measured $y(k)$ values, and $||\cdot||_2$ indicates a vector l_2-norm.

3.4.1 Data Pre-Processing and Model Structure Considerations

Data pre-processing tasks include interpolation (to account for missing data), mean subtraction, and shifting *Actual Steps* and *Granted Points* by one sample to reflect temporal precedence. Model structure selection decisions consist of determining, for each participant, the input signals to be included, and corresponding ARX model orders for the output and each input, in accordance with (3.4). Taking advantage of the computational simplicity associated with ARX modeling, the approach taken here is to exhaustively examine a range of model orders, and use model validation procedures to determine the most suitable structure. For this case study, ARX model order ranges for n_a and n_b from 1 to 3 (i.e., $\max(n_a) = 3$, and $\max(n_{b_j}) = 3 \quad \forall j = 1, \ldots, n_u$) seemed reasonable. *A priori* knowledge of the SCT fluid analogy model developed in Martín *et al.*, 2014 implies that very high order models should not be necessary to characterize these behavior-change dynamics. From inspecting the intervention data, it was reasonable to assume a basic unit input delay (i.e., $n_{k_j} = 1 \quad \forall j$).

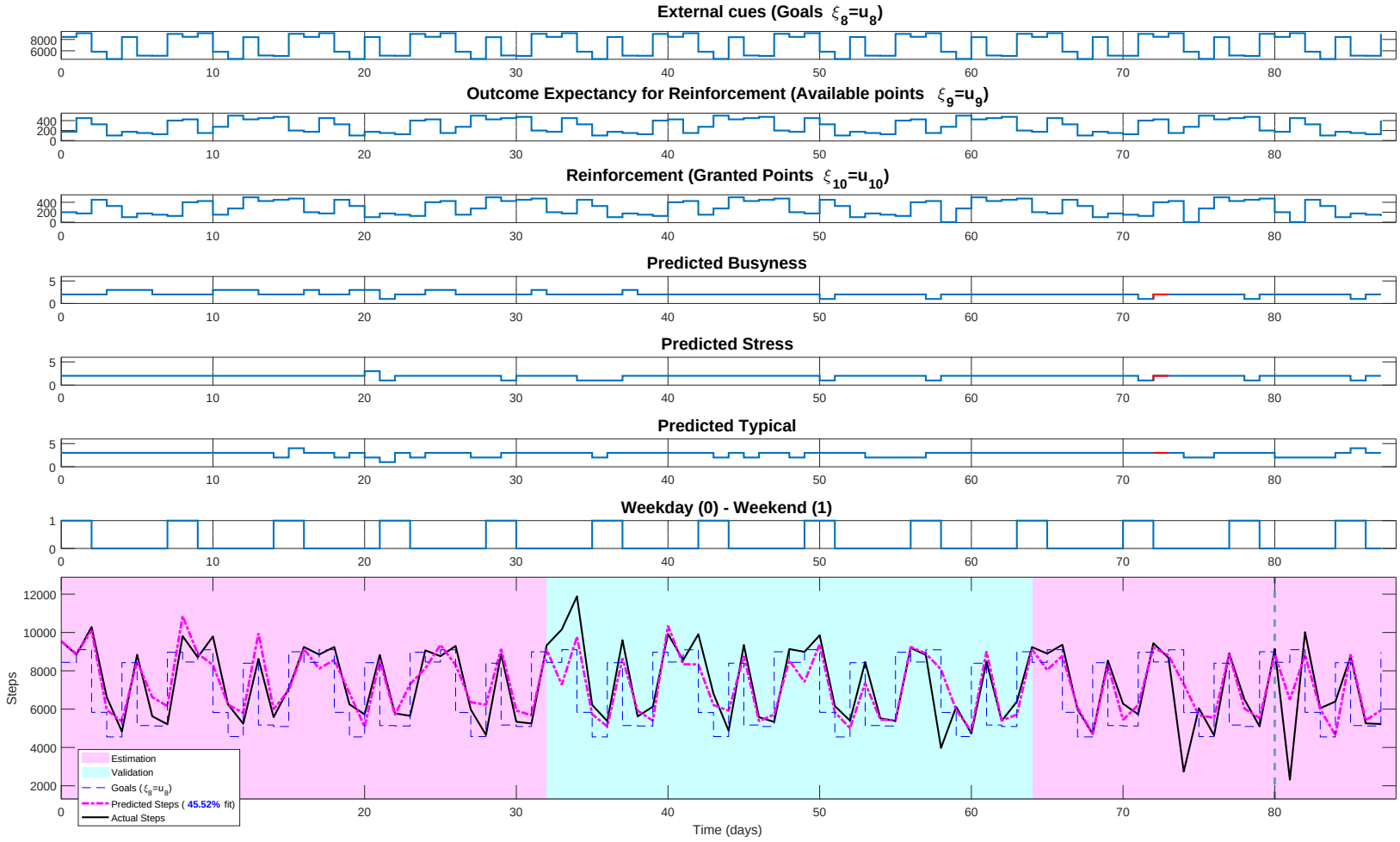

Figure 3.3: Time Series Plot Showing Seven Selected Input Sequences (Manipulated Inputs & Measured Disturbances), Predicted Behavior (from an ARX Black-box Model), Actual Behavior, Model Overall Fit, and Estimation & Validation Cycles (1st, 2nd, and 5th for Estimation; 3rd and 4th for Validation) for a Selected *Just Walk* Participant.

The absence of drifts and trends in the data leads to assume stationary (though potentially time-varying) noise characteristics over the course of the intervention period. In determining the inputs to be considered, the approach is to start with a basic 3-input model consisting of *Goals* (u_8), *Expected Points* (u_9), and *Granted Points* (u_{10}) and then add 4 additional measured inputs (*Predicted Busyness*, *Predicted Stress*, *Predicted Typical*, and *Weekday-Weekend*) to this basic model. All possible combinations of these inputs are estimated. Model validation following estimation ultimately determines which of these inputs are most important in describing individual behavior. Nonetheless, in the pre-processing stage, correlation analysis can be used to determine inputs that may be significantly crosscorrelated with each other or to identify inputs that appear to have no significant effect on the output. In both scenarios, the number of inputs that needs to be considered in the parameter estimation procedure can be reduced, ultimately leading to parsimonious models that can be generated with less effort.

3.4.2 Model Parameter Estimation and Validation

Model estimation and concomitant validation with the *Just Walk* intervention data is now considered. As mentioned earlier, first, a basic 3-input model was estimated and evaluated, followed by the addition and combination of 4 more inputs, leading to estimation of all possible combinations of these additional inputs. At an individual level, the full dataset was segmented into informative 16-day cycles for model estimation/validation. The cycle length was defined by the multisine input signal described in Section 3.3.

Cross-validation (the process of evaluating model fit over data not used for estimation) represents one of the most valuable aspects of system identification Ljung, 1994. The conventional approach in system identification is to assign a certain percentage

of data for estimation, followed by validation (e.g., 50% estimation, 50% validation). Such an approach assumes that the noise characteristics of the problem remain unchanged during the course of the intervention. However, it is reasonable to expect that noise and disturbance characteristics will vary over long-duration interventions such as *Just Walk*. In the analysis, each data cycle was assigned to either estimation or validation; all combinations of data cycles involving at least two cycles for validation were generated and evaluated.

Table 3.1 summarizes results of this procedure for a 4-input model (*Goals*, *Expected Points*, *Granted Points* and *Predicted Busyness*) of a selected participant. The fit index from Equation (3.5) was calculated for each cycle and averaged for estimation and validation data, respectively. All data cycle combinations that feature at least two cycles for validation or estimation (twenty candidate ARX models) were evaluated. For each of these combinations of estimation and validation cycles (corresponding to a specific row in Table 3.1), ARX orders were determined from an exhaustive search routine that selects a stable ARX model with highest predictive ability (based on the maximum average validation fit). This step provides a safeguard against overparametrization. The final chosen model should reflect, in addition to a good fit to validation data, a good fit for the entire data set (consisting of both estimation and validation cycles). This suggests that the final model choice should correspond to the model that yields highest overall fit (the "Overall NRMSE Fit" column in Table 3.1). Incorporating the overall fit criterion with the fit to cross-validation data balances good prediction with model accuracy over the entire data set. Note that using this analysis, the best results for the specific participant occur in the model resulting from row 18 (cycles 1, 2, and 5 for estimation; 3 and 4 for validation) with an overall NRMSE index at 46.03% for a model with structure $n_a = 2$, $n_{b_1} = 3$, $n_{b_2} = 1$, $n_{b_3} = 2$, and $n_{b_4} = 3$. This model performs close to the model with best fit

Table 3.1: Intermediate Results for a 4-input ARX Model of a Selected Participant from *Just Walk*

E*	V*	NRMSE Fit (%)					Average Estimation NRMSE Fit (%)	Average NRMSE Validation Fit (%)	Overall NRMSE Fit (%)	ARX Order (4-input) [na_1,nb_1,nb_2,nb_3,nb_4]
		Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5				
[1,2]	[3,4,5]	77.40%	85.44%	79.27%	27.68%	13.70%	81.42%	40.22%	40.11%	[1,1,1,1,3]
[1,3]	[2,4,5]	77.39%	82.25%	81.30%	26.88%	15.36%	79.35%	41.50%	40.60%	[1,2,1,1,3]
[1,4]	[2,3,5]	64.82%	71.25%	67.27%	45.89%	21.04%	55.36%	53.19%	42.29%	[1,3,1,1,1]
[1,5]	[2,3,4]	61.36%	59.51%	60.96%	40.14%	24.47%	42.92%	53.54%	37.40%	[1,1,1,3,1]
[2,3]	[1,4,5]	70.46%	90.25%	84.15%	25.00%	11.19%	87.20%	35.55%	37.70%	[3,3,1,2,3]
[2,4]	[1,3,5]	49.06%	71.94%	67.25%	52.39%	22.98%	62.17%	46.43%	40.56%	[3,1,2,1,3]
[2,5]	[1,3,4]	54.89%	61.75%	60.36%	47.35%	23.68%	42.72%	54.20%	39.33%	[3,1,1,1,1]
[3,4]	[1,2,5]	45.97%	61.27%	69.24%	51.46%	24.02%	60.35%	43.75%	41.15%	[1,3,3,1,3]
[3,5]	[1,2,4]	63.11%	66.96%	52.29%	41.52%	19.47%	35.88%	57.20%	41.12%	[1,1,1,1,1]
[4,5]	[1,2,3]	36.37%	52.47%	50.06%	49.24%	25.88%	37.56%	46.30%	32.75%	[1,1,1,3,2]
[3,4,5]	[1,2]	53.63%	64.61%	49.26%	46.59%	19.93%	38.59%	59.12%	40.12%	[1,1,1,1,1]
[2,4,5]	[1,3]	50.12%	59.76%	59.36%	49.92%	23.64%	44.44%	54.74%	38.71%	[3,1,1,1,1]
[2,3,5]	[1,4]	58.63%	66.76%	64.91%	49.62%	27.28%	52.98%	54.13%	40.59%	[3,1,3,2,1]
[2,3,4]	[1,5]	59.43%	76.99%	70.11%	41.51%	22.32%	62.87%	40.88%	41.61%	[2,3,3,2,3]
[1,4,5]	[2,3]	57.91%	61.11%	60.18%	45.69%	24.92%	42.84%	60.65%	38.81%	[1,1,1,3,1]
[1,3,5]	[2,4]	66.34%	66.02%	67.24%	42.13%	22.57%	52.05%	54.08%	41.31%	[1,3,1,1,1]
[1,3,4]	[2,5]	68.39%	77.75%	73.46%	41.86%	18.78%	61.24%	48.27%	42.26%	[1,3,2,1,1]
[1,2,5]	[3,4]	61.85%	56.05%	68.43%	44.82%	35.02%	50.97%	56.63%	46.03%	[2,3,1,2,3]
[1,2,4]	[3,5]	71.99%	73.18%	72.36%	43.28%	20.40%	62.82%	46.38%	43.61%	[1,2,1,1,3]
[1,2,3]	[4,5]	75.95%	87.02%	80.67%	26.39%	13.36%	81.21%	19.88%	39.87%	[1,1,1,1,3]

*E≡Estimation Cycles (magenta), V≡Validation Cycles (cyan)

over the validation data (average validation fit of 56.63% for row 18 vs 60.65% in row 15); however, the model with the best fit to validation data does not exhibit the best fit to data overall (38.81% in lieu of 46.03%).

3.4.3 Overall Fit Analysis and Assessment of Individual Participant Characteristics

Similar analyses to those presented in Table 3.1 can be performed with additional inputs, for all possible combinations. For example, for a total of 7 inputs, 16 different input models can be generated for each participant (since *Goals* (u_8), *Expected Points* (u_9), and *Granted Points* (u_{10}) are always grouped). Evaluating these 16 input combinations allows us to draw conclusions on participant characteristics that resulted from the intervention.

Figure 3.4 depicts model validation % fit results from three different participants from *Just Walk*. The Y-axis indicates the % fit of the 3, 4, 5, 6, and 7-input models,

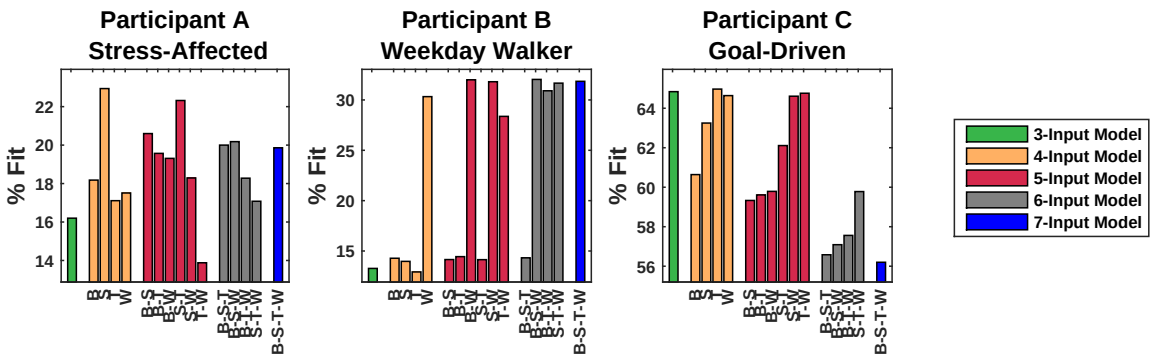

Figure 3.4: Average Validation % Fits of Individualized ARX Models from Black-box System Identification for Three Individuals: Goals-Expected Points-Granted Points Model (Base Inputs); **B**: Predicted Busyness; **S**: Predicted Stress; **T**: Predicted Typical; **W**: Weekday-Weekend.

and the X-axis corresponds to the psychosocial measures (busyness, stress, weekday, typical) measured daily. Here, it is seen that Participant A's walking behavior is largely driven by stress (highest % fit seen for the stress bar in the 4-input model), Participant B's behavior is driven by whether it is a weekday or weekend, while Participant C has the highest % fit for the 3-input model, indicating that the daily step goal had the greatest impact on walking behavior. Step responses from the individual ARX models can be used to reveal more precise directionality and magnitude information; for example, from Figure 3.5, one can predict that the selected participant will typically reach 80% of the desired daily step goals within the first day of goal announcement. Responsiveness to other inputs and disturbances can be determined similarly. This strategy has significant implications for personalized and adaptive behavior change interventions; if one can determine the inputs that are most meaningful for a given individual in a given context, it is possible then to optimize the target behavior over a specified time (hours, days, weeks, months).

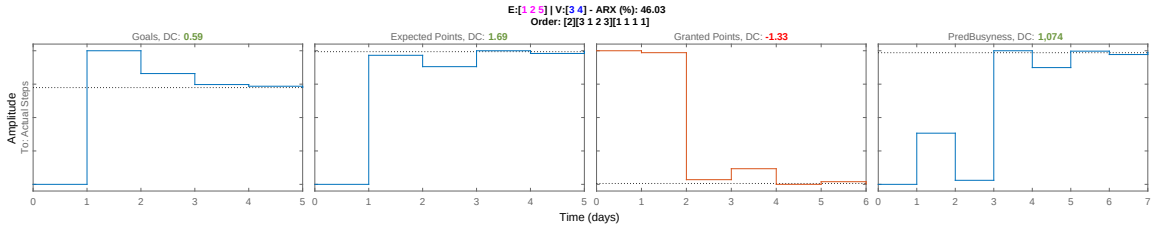

Figure 3.5: Step Responses of a 4-input Model (Including *Predicted Busyness*) for a Selected Participant.

3.5 Chapter Summary

This chapter presented a system identification modeling strategy for a physical activity (e.g., walking) intervention delivered via a smartphone application. The results from this dataset represent an important accomplishment in understanding behavior-change from a data-driven perspective. Predictive and consistent black-box models are crucial for validating behavioral theory (such as the SCT model). It is shown that segmenting and evaluating the data at a per-cycle level gives the most valid results to date. These types of models are necessary to effectively model behavior, which is highly complex, idiosyncratic, and dynamic in nature. In addition, drawn from the analysis of the estimated models, it is important in experimental design to consider capturing more low-frequency dynamics (e.g., using pseudo-random binary sequences), to draw more decisive conclusions on participant long-term (steady-state) responses. Finally, the enhanced identification testing monitoring procedure in Martín *et al.*, 2016b can be considered in future experiment design.

Chapter 4

A SPECTRAL DECOMPOSITION IDENTIFICATION FORMULATION FOR STRUCTURED STATE-SPACE MODELS: ESTIMATING SEMIPHYSICAL MODELS OF SOCIAL COGNITIVE THEORY

4.1 Background

The *grey-box* identification problem, i.e., the estimation of physical/semiphysical state-space models with an imposed structure, is a topic of increasing interest and a growing number of applications. For physical, semiphysical, and compartmental/network models, the resulting state-space realizations often emerge in continuous-time with some sort of a structure. Furthermore, in the vast majority of structured state-space models, an imposed or emerged model structure is often linear in the sense that some entries of the system matrices are determined *a priori*, and can be expressed as a set of linear equations (e.g., $a_{ij} = 0$, $a_{ij} + c_1 b_{11} = c_2$, $b_{ij} = c_3$ for some indices i and j; $c_{1,2,3} \in \mathbb{R}$ are some constants). We refer to such models as *linearly-structured*. A subset of this class of structures has been presented as *affinely parametrized structures* (Yu *et al.*, 2019). As expressed in Ljung, 2019, considering linear structures is not very restrictive; most of structured models belong either readily or after a reformulation. Motivated by the real-life behavioral experiment *Just Walk*, and the need for data-driven semiphysical models of Social Cognitive Theory (SCT), we present a spectral decomposition (SD) identification algorithm for estimating this specific class of models, i.e., *linearly-structured* models.

While it is recognized that the general prediction-error method (PEM), which estimates the constrained model directly from Input/Output (I/O) data, has the best

possible asymptotic properties Yu *et al.*, 2018, it is also known that due to the non-convexity of the problem, "the domain of attraction of the global minimum is not very forgiving" (Ljung, 2019), and the method is thus crucially dependent on good initial estimates for its practical success in problems of realistic sizes. This particular difficulty has been illustrated by the numerical experiment presented in Ljung, 2019 from Parrilo and Ljung, 2003 and will be further illustrated with an additional numerical example in this chapter (see Section 4.3.2). Indeed, this solver initialization challenge is amplified when the structure in question is inherently unidentifiable, leading to the fact that even if a global minimum is found, it need not be unique. In fact, in this case, an infinity of global minima solutions may be on offer, and hence the physical significance of model parameters need not be preserved. Note that, here, model *identifiability* is defined per, Ljung, 1999, Definition 4.6. Nevertheless, in this chapter, we present a formulation that aims at eliminating or at least significantly reducing the burden on the user with the challenge of judicious solver initialization, particularly in the absence of sufficient prior knowledge. This may be possible at the expense of increased computational load while solving multiple random initializations/restarts that may be needed in harder problems.

The notion of initializing PEM for grey-box model estimation from black-box models that are fully-parameterized (i.e., "total models" in Ljung, 2019) has been suggested and standardized in numerous settings before; the reader is referred to Yu *et al.*, 2019; Yu *et al.*, 2018; Parrilo and Ljung, 2003; Mercère *et al.*, 2014 and references therein for elaboration. In particular, subspace identification methods (e.g., N4SID, MOESP) seemed favorable given their asymptotic consistency (Yu *et al.*, 2018; Parrilo and Ljung, 2003). Similarly, building from an identified black-box model, the main premise of the proposed SD formulation is to exploit the known linear structure and formulate an easier similarity transformations search, restricting the search space to

only of that which is fundamentally unknowable *a priori*.

The rest of this chapter is organized as follows: Section 4.2 presents the SD formulation for estimating linearly-structured, continuous-time state-space models. Sufficient conditions for the existence and uniqueness of SD models is presented with additional conclusions on identifiability. Next, Section 4.3 provides two numerical examples and introduces a further simplified SCT model, followed by a brief report on results from *Just Walk* in Section 4.4. Finally, conclusions and potential future directions are summarized in Section 4.5.

4.2 Spectral Decomposition Formulation

Using measured input-output sequences $\{U_k, Y_k\}_{k=1}^{N}$, consider a minimal, fully-parametrized discrete-time model $\hat{\mathcal{S}}_d(\hat{A}_d, \hat{B}_d, \hat{C}_d, \hat{D}_d, \hat{K}_d)$ estimated using a classical subspace method (i.e., N4SID, Van Overschee and De Moor, 2012), satisfying the following set of linear equations

$$\begin{bmatrix} \hat{X}_{k+1} \\ Y_k \end{bmatrix} = \begin{bmatrix} \hat{A}_d & \hat{B}_d \\ \hat{C}_d & \hat{D}_d \end{bmatrix} \begin{bmatrix} \hat{X}_k \\ U_k \end{bmatrix} + \begin{bmatrix} \hat{K}_d \\ I \end{bmatrix} \mathcal{E}_k \tag{4.1}$$

where $\hat{X}_k \in \mathbb{R}^{n \times N}$, $Y_k \in \mathbb{R}^{p \times N}$, and $\mathcal{E}_k \in \mathbb{R}^{p \times N}$ is a column-wise sequence of noise in innovations form (i.e., $\mathcal{E} \sim \mathcal{N}(0, \sigma_i)\ \forall i \in \{1, \ldots, p\}$). With respect to I/O data $\{U_k, Y_k\}$, the identified states from the subspace method are modulo a similarity transformation T_f (i.e., $\hat{X}_k = T_f X_k$) with $\hat{\mathcal{S}}_d(T_f A_d T_f^{-1}, T_f B_d, C_d T_f^{-1}, D_d, T_f K_d)$. For the identified system $\hat{S}_d$ of McMillan degree n, m inputs, and p measured outputs, the minimal realization system matrices are $A_d \in \mathbb{R}^{n \times n}$, $B_d \in \mathbb{R}^{n \times m}$, $C_d \in \mathbb{R}^{p \times n}$, $D_d \in \mathbb{R}^{p \times m}$, $K_d \in \mathbb{R}^{n \times p}$, and $T_f \in \mathbb{R}^{n \times n}$. For simplicity, and without loss of generality, we proceed in the various parts of this chapter with a strictly proper, disturbance-free model of $\hat{\mathcal{S}}_d(T_f A_d T_f^{-1}, T_f B_d, C_d T_f^{-1})$.

4.2.1 Main Formulation

Through the remainder of this chapter, two main assumptions follow: First, we assume that $\hat{A}_d$ is of n distinct eigenvalues. Second, we assume Zero-Order Hold (ZOH) intersample behavior. The ZOH assumption becomes more practical in experiments with smaller sampling times (i.e., $T_s \to 0$). Benefiting from assuming ZOH, we recall that the continuous-time image of $\hat{\mathcal{S}}_d$, $\hat{\mathcal{S}}(A, B, C)$, can be obtained from the well-known exact discretization equations from linear systems theory

$$A_d = e^{AT_s} \tag{4.2a}$$

$$B_d = \int_0^{T_s} e^{A\tau} B u(\tau) d\tau \overset{\text{ZOH}}{=} A^{-1} \left(A_d - I_n\right) B \tag{4.2b}$$

$$C_d = C \tag{4.2c}$$

where T_s is the sampling time. By assuming n distinct poles of the identified $\hat{\mathcal{S}}_d$, we allow for the following spectral decomposition

$$A_d = T \Lambda_d T^{-1} \Rightarrow A = T \Lambda T^{-1}, \quad \Lambda = \log(\Lambda_d)/T_s \tag{4.3}$$

Thus, an estimate for the state-transition matrix subject to a linear structure imposed by the pair $(P_A, d_A) := \{(P_A, d_A) : \ P_A \theta_A = d_A, \ \theta_A = \text{vec}\{A\}\}$ is obtained by solving the linearly-constrained eigenvalue problem

$$T^{-1} A T = \Lambda \tag{4.4}$$

subject to the structure pair (P_A, d_A), from which the following quadratic program (QP) follows

$$\begin{aligned} \min_{\theta_A} \quad & \frac{1}{2} \theta_A^T H_A \theta_A + f_A^T \theta_A \\ \text{s.t.} \quad & P_A \theta_A = d_A \end{aligned} \tag{4.5}$$

where

$$
\begin{aligned}
H_A &= 2(T^T \otimes T^{-1})^T (T^T \otimes T^{-1}) \\
f_A &= \left[-2\varphi_A^T (T^T \otimes T^{-1})\right]^T \\
\varphi_A &= \mathrm{vec}\{\Lambda\}
\end{aligned}
$$

and $\otimes$ denoting the Kronecker product. Similarly, by assuming ZOH, one can use (4.2b) and (4.3) to establish

$$
T^{-1}B = (\Lambda_d - I_n)^{-1} \Lambda T^{-1} B_d \tag{4.7}
$$

for estimating the input gain matrix B subject to the structure (P_B, d_B), from which it also follows

$$
\begin{aligned}
\min_{\theta_B} \quad & \frac{1}{2}\theta_B^T H_B \theta_B + f_B^T \theta_B \\
\text{s.t.} \quad & P_B \theta_B = d_B
\end{aligned} \tag{4.8}
$$

with

$$
\begin{aligned}
H_B &= 2\left(I_m \otimes T^{-1}\right)^T \left(I_m \otimes T^{-1}\right) \\
f_B &= -2\left[\varphi_B^T \left(I_m \otimes T^{-1}\right)\right]^T \\
\varphi_B &= \mathrm{vec}\{\Gamma^{-1} T_0^{-1} B_d^0\} \\
\Gamma &= \Lambda^{-1}\left(\Lambda_d - I_n\right)
\end{aligned}
$$

Note that the *characteristic* matrix Γ is system-invariant (i.e., Γ is unique and does not depend on the applied similarity transformation); this extends the invariance to φ_B (similar to φ_A) once $\hat{S}_d^0$ of A_d^0 (thereby T_0) and B_d^0 is initially established. By defining the Kronecker eigenvector matrix

$$
\mathcal{T} \stackrel{\text{def}}{=} \begin{bmatrix} T \otimes T^{-T} & \mathbf{0}_{n^2 \times nm} \\ \mathbf{0}_{nm \times n^2} & I_m \otimes T^{-T} \end{bmatrix} \tag{4.10}
$$

and using the Kronecker product properties, it can be shown that in a more compact form, (4.5) and (4.8) can be solved simultaneously by the following augmentation

$$\begin{aligned} \min_{\theta} \quad & \frac{1}{2}\theta^T H \theta + f^T \theta \\ \text{s.t.} \quad & P\theta = d \end{aligned} \tag{4.11}$$

with

$$\begin{aligned} H &= 2\begin{bmatrix} TT^T \otimes \left(TT^T\right)^{-1} & \mathbf{0}_{n^2 \times nm} \\ \mathbf{0}_{nm \times n^2} & I_m \otimes \left(TT^T\right)^{-1} \end{bmatrix} = 2\mathcal{T}\mathcal{T}^T \\ f &= -2\begin{bmatrix} T \otimes T^{-T} & \mathbf{0}_{n^2 \times nm} \\ \mathbf{0}_{nm \times n^2} & I_m \otimes T^{-T} \end{bmatrix} \varphi = -2\mathcal{T}\varphi \\ \varphi &= \begin{bmatrix} \varphi_A^T & \varphi_B^T \end{bmatrix}^T \\ \theta &= \begin{bmatrix} \theta_A^T & \theta_B^T \end{bmatrix}^T \\ P &= \begin{bmatrix} P_A & \mathbf{0}_{p_a \times nm} \\ \mathbf{0}_{p_b \times n^2} & P_B \end{bmatrix} \\ d &= \begin{bmatrix} d_A^T & d_B^T \end{bmatrix}^T \end{aligned}$$

where $\circ^{-T} = \left(\circ^T\right)^{-1} = \left(\circ^{-1}\right)^T$. Note that in cases where some linear relations exist between parameters of the structured A and B, the augmented QP in (4.11) is used. However, the pairs (P_A, d_A) and (P_B, d_B) would no longer separable, but constructed as a single (P, d) instead; P will no longer be block-diagonal.

4.2.2 Existence, Uniqueness, and Identifiability

With the main formulation delivered in (4.11), we now proceed to establish the existence and uniqueness of an SD-identified model, as well as present a corollary on identifiability.

Theorem 4.2.1 (Existence)**.** *For the minimal discrete-time model $\mathcal{S}_d$ in* (4.1) *with n distinct eigenvalues of A_d and ZOH intersample behavior on all inputs, a sufficient condition for the existence of $\mathcal{S}_{[P,d]}$ with a linearly-structured θ^*_{SD} subject to (P, d) in* (4.11) *is $|PH^{-1}P^T| \neq 0$.*

Proof. The explicit solution of (4.11) is known (Boyd and Vandenberghe, 2004)

$$\begin{bmatrix} \theta^*_{SD} \\ \lambda^* \end{bmatrix} = \begin{bmatrix} H & P^T \\ P & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} -f \\ d \end{bmatrix} \tag{4.13}$$

and exists for some Lagrange multipliers λ^* if the Karush-Kuhn-Tucker (KKT) matrix is nonsingular. Using block matrix inversion, i.e.,

$$\begin{bmatrix} H & P^T \\ P & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} H^{-1} - H^{-1}P^T(PH^{-1}P^T)^{-1}PH^{-1} & H^{-1}P^T(PH^{-1}P^T)^{-1} \\ (PH^{-1}P^T)^{-1}PH^{-1} & -(PH^{-1}P^T)^{-1} \end{bmatrix} \tag{4.14}$$

yields

$$\theta^*_{SD} = H^{-1}P^T\Psi^{-1}d + \left(H^{-1}P^T\Psi^{-1}P - I\right)H^{-1}f \tag{4.15}$$

where $\Psi = PH^{-1}P^T$. A_d of n distinct eigenvalues implies T is nonsingular, which establishes the existence of H^{-1}. Hence, a sufficient condition for the existence of θ^*_{SD} (thereby Ψ^{-1}) is $|PH^{-1}P^T| \neq 0$, i.e., the imposed structure P is feasible. □

Note that $H = 2\mathcal{T}\mathcal{T}^T$ in (4.11) implies that this QP is convex. Further, by extension, if the *quadratic* structural constraints in (6.1) are convex, the resulting QCQP remains convex.

Theorem 4.2.2 (Uniqueness)**.** *For the minimal discrete-time model $\mathcal{S}_d$ in* (4.1) *of order n and p outputs, the linearly-structured, SD-identified continuous-time model $\mathcal{S}_{[P,d]}$ subject to (P, d) in* (4.11) *such that $|PH^{-1}P^T| \neq 0$ is unique when $p = n$, $|C| \neq 0$.*

Proof. Given $|PH^{-1}P^T| \neq 0$ establishes the nonsingularity of the KKT matrix in (4.14) and implies that $[H^T \ P^T]^T$ has linearly independent columns. It follows that θ^*_{SD} of (4.11) is unique; full proof in Boyd and Vandenberghe, 2004. Hence, a sufficient condition for the uniqueness of the SD model follows from the uniqueness of the range space of T in $\mathbb{R}^n$, $\mathcal{R}(T)$, which is achieved when $p = n$, $|C| \neq 0$; see, Bellman and Åström, 1970, Sec. 6 and, Delforge, 1982, Sec. 3. □

Corollary 4.2.2.1 (Identifiability)**.** *For a minimal discrete-time model $\mathcal{S}_d$ of order n and p outputs, assume ZOH intersampling behavior on all inputs, and define the set of similarity transformations $\mathscr{T}_f$*

$$\mathscr{T}_f := \left\{ T_f : \; T_f = \begin{bmatrix} I_p & & \mathbf{0}_{n-p} \\ - & H_f & - \end{bmatrix}, \; H_f \in \mathbb{R}^{(n-p)\times n}, \; |T_f| \neq 0 \right\}$$

The identifiability of the continuous-time $\mathcal{S}$ is as follows:

- *$p = n$, $|C| \neq 0$, and $|PH^{-1}P^T| \neq 0$: the continuous-time $\mathcal{S}_{[P,d]}$ subject to (P, d) is globally identifiable; the estimator* (4.15) *yields a unique estimate for all system matrices.*

- *$p < n$, $C = (\, I_p \; \mathbf{0}_{n-p} \,)$: at least the first p rows of B and K are uniquely identifiable, modulo $\mathcal{E}_k$, in the fully-parametrized $\mathcal{S}$ (i.e., $(P, d) = \{\varnothing\}$) $\forall \; T_f \in \mathscr{T}_f$.*

Proof. When $p = n$, the SD-identified model is unique, and the proof follows immediately from Theorem 4.2.2. When $p < n$, recall that from (4.1) we have

$$\begin{aligned} \hat{X}_{k+1} &= T_f A_d T_f^{-1} \hat{X}_k + T_f B_d U_k + T_f K_d \mathcal{E}_k \\ Y_k &= C T_f^{-1} \hat{X}_k + D_d U_k + \mathcal{E}_k \end{aligned}$$

which by substitution gives

$$Y_k = C A_d T_f^{-1} \hat{X}_{k-1} + C B_d U_{k-1} + C K_d \mathcal{E}_{k-1} + D_d U_k + \mathcal{E}_k \tag{4.16}$$

With $C = (\, I_p \; \mathbf{0}_{n-p} \,)$, (4.16) shows that entries of the first p rows of B_d and K_d must be unique $\forall \; T_f \in \mathscr{T}_f$; noting that $B = T_f T \Gamma^{-1} T^{-1} B_d$, $T_f \in \mathscr{T}_f \implies T_f^{-1} \in \mathscr{T}_f$ by block inversion, and the augmentation $B_d' = [B_d \; K_d]$ and $U_k' = [U_k^T \; \mathcal{E}_k^T]^T$ extends the result to B and K. □

4.2.3 Model Realization: An LMI Approach

In the identification of structured state-space models of physical and other types of systems, the model order is often determined *a priori* and the model must be physically realizable. Hence, when opting in for SD-identified models, we know from (4.3) that Λ_d must at least be positive semidefinite. However, in many cases when estimating discrete-time models, the presence of process/measurement noise or unmodeled dynamics may suffice to explain the emergence of negative eigenvalues, i.e., physically unrealizable models of ZOH inversion. To circumvent this potential obstacle for the SD user, we include a brief review on the work of Miller *et al.*, 2012; Miller and De Callafon, 2013 that ensures $\Lambda_d \succeq 0$ by applying Linear Matrix Inequalities (LMI) constraints in the earlier step of estimating fully-parametrized subspace models in (4.1).

Per Miller and De Callafon, 2013, the main theorem in Chilali and Gahinet, 1996 establishes that the eigenvalues of a matrix A_d lie within an LMI region defined by the characteristic function

$$f_{\mathcal{D}}(z) = \alpha + \beta z + \beta^T \bar{z}$$

with

$$\mathcal{D} = \{z \in \mathbb{C} : \ f_{\mathcal{D}}(z) \geq 0\}$$

if and only if $\exists\ P_d \in \mathbb{R}^{n \times n}$ such that

$$P_d = P_d^T \succ 0, \qquad \mathcal{M}_{\mathcal{D}}(Q_d, P_d) \succeq 0$$

where

$$\mathcal{M}_{\mathcal{D}}(Q_d, P_d) = \alpha \otimes P_d + \beta \otimes Q_d + \beta^T \otimes Q_d^T$$

with α (symmetric) and β (square) are generally not unique matrices establishing the feasible LMI region $\mathcal{D}$ in the z-plane. Given matrices R_1, R_2, and W, depending on the underlying subspace method, the new eigenvalue-constrained A_d^* can then be

found by solving the following convex, semidefinite program

$$\begin{aligned} \min_{Q_d, P_d} \quad & \|R_1 Q_d - W R_2^{\dagger} P_d\|_F \\ \text{s.t.} \quad & \mathcal{M}_{\mathcal{D}}(Q_d, P_d) \succeq 0 \\ & P_d = P_d^T \succ 0 \\ & \operatorname{tr}(P_d) = l \end{aligned} \tag{4.17}$$

where $Q_d \triangleq A_d P_d$, $l > 0$ is a positive scalar ($l = 1$ in Miller *et al.*, 2012) that ensures P_d does not become arbitrarily small during the minimization procedure; $\dagger$ and $\|\cdot\|_F$ denoting the Moore-Penrose pseudoinverse and the Frobenius norm, respectively. The eigenvalue-constrained A_d^* is then found as $A_d^* = Q_d^* P_d^{*-1}$. When the used subspace method is the deterministic N4SID, R_1, R_2, and W are defined below:

$$R_1 = I_n, \quad R_2 = \hat{X}_k, \quad W = \hat{X}_{k+1} - B_d U_k \tag{4.18}$$

For additional subspace methods, including the stochastic N4SID, corresponding values for R_1, R_2, and W are tabulated in Miller *et al.*, 2012.

4.2.4 Standard SD Loss Function

As noted in Section 4.2, if $\hat{C}_d$ in (4.1) is full-rank, the SD model is unique and obtained analytically by evaluating (4.15). Otherwise, a search for a suitable similarity transformation may be in order. Consider

$$J_0 = \left(\mathcal{Y} - \Phi_0 \theta_d^0\right)^T \left(\mathcal{Y} - \Phi_0 \theta_d^0\right) \tag{4.19}$$

where $\Phi_0 = \left[X_{0,k-1}^T \; U_{k-1}^T\right] \otimes C_0$; $\mathcal{Y}$ and θ_d^0 denote the vectorized measured outputs and the initial discrete-time parameter vector, respectively (note the typological error in Freigoun *et al.*, 2021b). C_0 is structurally feasible and is estimated in discrete-time (e.g., using PEM) if unknown *a priori*. When applying a similarity transform T_f,

mapping from $\mathcal{S}_d^0$ to $\mathcal{S}_d^1$ (*both* belonging to the set of structurally feasible models), it can be shown that

$$\Phi_1 = \Phi_0 \mathcal{T}_f^T, \quad \theta_d^1 = \mathcal{T}_f^{-T} \theta_d^0 \overset{\text{ZOH}}{=} \mathfrak{T}_f \Gamma \mathfrak{T}_f^{-1} \theta_{SD}^1$$

where

$$\mathfrak{T}_f = \begin{bmatrix} (T_f T_0)^{-T} \otimes I_n & \mathbf{0}_{n^2 \times nm} \\ \mathbf{0}_{nm \times n^2} & I_m \otimes T_f T_0 \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \Lambda_d \Lambda^{-1} \otimes I_n & \mathbf{0}_{n^2 \times nm} \\ \mathbf{0}_{nm \times n^2} & I_m \otimes \Gamma \end{bmatrix} \tag{4.20}$$

and T_0 is a well-conditioned initial eigenvector matrix in (4.3). By using (4.15), θ_{SD} is obtained in terms of T_f as follows

$$\begin{aligned} \theta_{SD} = {} & \mathcal{T}_f^{-T} H_0^{-1} \mathcal{T}_f^{-1} P^T \left(P \mathcal{T}_f^{-T} H_0^{-1} \mathcal{T}_f^{-1} P^T \right)^{-1} d \\ & - \mathcal{T}_f^{-T} H_0^{-1} \mathcal{T}_f^{-1} P^T \left(P \mathcal{T}_f^{-T} H_0^{-1} \mathcal{T}_f^{-1} P^T \right)^{-1} P \mathcal{T}_f^{-T} \mathcal{T}_0^{-T} \varphi \\ & + \mathcal{T}_f^{-T} \mathcal{T}_0^{-T} \varphi \end{aligned} \tag{4.21}$$

where $H_0^{-1} = 0.5 \mathcal{T}_0^{-T} \mathcal{T}_0^{-1}$. Finally, with $\hat{\Phi}(T_f) \overset{\text{def}}{=} \Phi_0 \mathcal{T}_f^T \mathfrak{T}_f \Gamma \mathfrak{T}_f^{-1}$, the standard SD loss function J_{SD} is minimized with respect to T_f, i.e.,

$$\begin{aligned} \min_{T_f} \quad & \left(\mathcal{Y} - \hat{\Phi} \theta_{SD} \right)^T \left(\mathcal{Y} - \hat{\Phi} \theta_{SD} \right) \\ \text{s.t.} \quad & C_0 T_f = C_0 \end{aligned} \tag{4.22}$$

When $\mathcal{Y}$ is replaced with $\mathcal{Y}_{\text{sim}} = \Phi_0 \theta_d^0$, minimizing (4.22) translates into a more direct model-matching problem. It is recognized that the objective function in (4.22) describes an *approximate* estimate of the actual simulation error; the appeal is in its quasi-linear form (hence referred to as "standard"). The SD formulation can, however, be applied to the standard PEM formulation for an exact description of the simulation error during the minimization procedure. Furthermore, while also nonlinear, in more practical settings it may be interesting to also consider applying a direct NRMSE objective in 3.5 jointly with (4.21) in lieu of (4.22) and classical PEM to arrive at an SD model for PEM initialization.

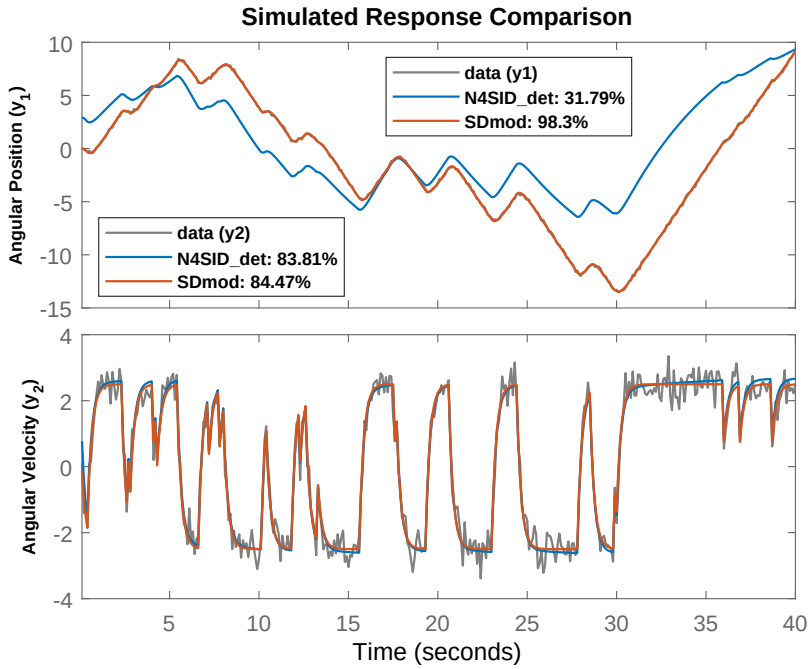

Figure 4.1: Simulation of the SD-identified DC Motor Model (Continuous-Time) From (4.15) Contrasted Against the Deterministic N4SID Model.

4.3 Numerical Examples

4.3.1 DC Servo Motor ($p = n$, $|C| \neq 0$)

In this illustration, we borrow the classical DC servo motor example presented in Ljung, 1999 by loading the demo **`iddemo7`** of the **`ident`** toolbox in MATLAB® Online R2021a. While this is a very basic example with near ideal conditions, it serves as a suitable illustration for using (4.15) since, here, a full-rank C is given and both assumptions of Theorem 4.2.1 are satisfied. The simulation of the deterministic N4SID model (with $\Lambda_d \succeq 0$) is shown in Figure 4.1 (blue: 31.79% & 83.81% NRMSE fits). Given the physical model

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & \theta_1 \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ \theta_2 \end{bmatrix} u(t) \tag{4.23a}$$

$$y(t) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x(t) + e(t) \tag{4.23b}$$

one can observe a slight dislocation of one eigenvalue from the origin to -0.1578. To calculate the structured SD model, the pair (P, d) is constructed from (4.23a)

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}, \quad d = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \tag{4.24}$$

and the solution is obtained analytically from (4.15) in a single step; simulations are in Figure 4.1 (red: 98.3% & 84.47%). This improvement can only be explained by the use of the structural information known *a priori* to correct for fitting to noise. It is also noted that the estimated SD model from the stochastic N4SID (using the `n4sid` command in MATLAB®) gives slightly improved results (98.32% & 84.47%) that level with the reported PEM solution (98.35% & 84.43%).

4.3.2 OCSE Fluid Analogy Model of Social Cognitive Theory

The purpose of this example is to present a numerical experiment on a newly introduced, further simplified fluid analogy model of Social Cognitive Theory Martín *et al.*, 2014; Freigoun *et al.*, 2017; the Operant Conditioning-Self-Efficacy (OCSE) loops model in Figure 4.2. A visualization of the results is provided to illustrate the partial identifiability case of Corollary 4.2.2.1. In Section 4.4, we use the OCSE structure in this example to estimate individual semiphysical SCT models using input-output participant data retrieved from the *Just Walk* study (see Appendix A). Using the conservation of mass principle, the OCSE model in Figure 4.2 can be represented by the following state-space model

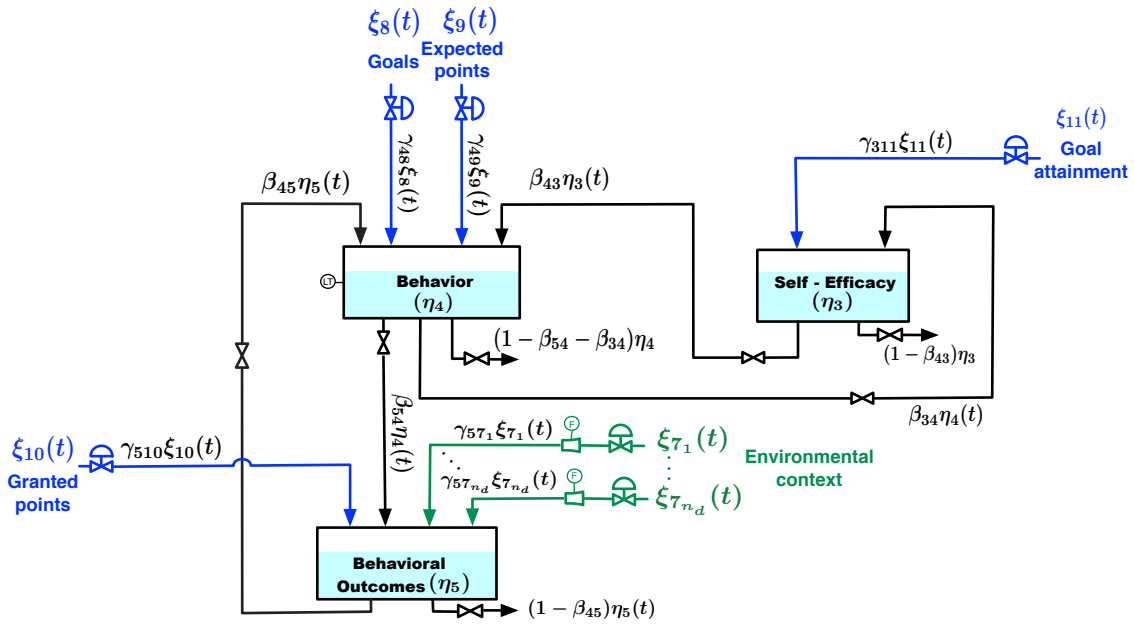

Figure 4.2: Further Simplified Fluid Analogy of Social Cognitive Theory (OCSE Model).

$$\dot{x}(t) = A(\theta_p)x(t) + B(\theta_p)u(t) \tag{4.25a}$$
$$y(t) = Cx(t) + e(t) \tag{4.25b}$$

where $x(t) = [\eta_4(t) \quad \eta_3(t) \quad \eta_5(t)]^T \in \mathbb{R}^3$,
$u(t) = \left[\xi_8(t) \quad \xi_9(t) \quad \xi_{10}(t) \quad \xi_{7_1}(t) \quad \dots \quad \xi_{7_{n_d}}(t)\right]^T \in \mathbb{R}^{3+n_d}$, $y(t) = \eta_4(t) \in \mathbb{R}$. With $\xi_{11}(t) \triangleq \eta_4(t) - \xi_8(t)$, we have

$$A(\theta_p) = \begin{bmatrix} -1/\tau_4 & \beta_{43}/\tau_4 & \beta_{45}/\tau_4 \\ (\gamma_{311}+\beta_{34})/\tau_3 & -1/\tau_3 & 0 \\ \beta_{54}/\tau_5 & 0 & -1/\tau_5 \end{bmatrix}$$

$$B(\theta_p) = \begin{bmatrix} \gamma_{48}/\tau_4 & \gamma_{49}/\tau_4 & 0 & 0 & \dots & 0 \\ -\gamma_{311}/\tau_3 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \gamma_{510}/\tau_5 & \gamma_{7_1}/\tau_5 & \dots & \gamma_{7_5}/\tau_5 \end{bmatrix}$$

$$C = [1 \quad 0 \quad 0]$$

with $n_d = 5$, the parameter vector $\theta_p \in \mathbb{R}^{16}$ is defined below

$$\theta_p = [\tau_4,\ \tau_3,\ \tau_5,\ \beta_{43},\ \beta_{45},\ \beta_{34},\ \beta_{54},\ \gamma_{48},\ \gamma_{49},\ \gamma_{311},\ \gamma_{510},\ \gamma_{7_1},\ \gamma_{7_2},\ \gamma_{7_3},\ \gamma_{7_4},\ \gamma_{7_5}]^T$$

In this example, 83 output sequences from the model in (4.25) with the hypothetical parameters

$$\theta_p = [0.2,\ 4,\ 0.5,\ 0.25,\ 0.25,\ 0.25,\ 0.25,\ 0.75,\ 0.5,\ 3,\ -5,\ 350,\ 400,\ 300,\ -50,\ -250]^T$$

were generated by applying a subset of input-output data from a *Just Walk* Participant (published data were digitized according to the procedure detailed in Appendix A); see Freigoun *et al.*, 2017; Mercere, 2017; Phatak *et al.*, 2018. Next, similar to the actual implementation of *Just Walk* study, where only $\eta_4(t)$ is measured, the hypothetically generated outputs sequences $\eta_3(t)$ and $\eta_5(t)$ were discarded. An additional input was constructed (similar to Weekday/Weekend) for the purposes of this example. Following Corollary 4.2.2.1 and assuming no model mismatch (here, we know this is the case), one expects to recover the true first row of $B(\theta_p)$, including $\gamma_{48}/\tau_4 = 3.75$ and $\gamma_{49}/\tau_4 = 2.5$, as $e(t) \to 0$. This notion of partial identifiability is best visualized in Figure 4.3, where the poor initialization (intentional)

$$A(\theta_{p_0}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad B(\theta_{p_0}) = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

was used to obtain the classical PEM models (**`ssest`** in MATLAB) in the first column. For each SNR scenario, 1000 different zero mean Gaussian noise realizations were made to corrupt the 'true' output; a random $T_f \in \mathscr{T}_f$ is the applied to the full N4SID model with every noise realization. The first column shows results from PEM when poorly initialized; the second column shows estimates from the N4SID method; the third column shows PEM estimates when initialized by the N4SID model in the second column. The first two rows feature two identifiable parameters in $B(\theta_p)$ per Corollary 4.2.2.1; the third row features an unidentifiable parameter in $B(\theta_p)$. Results of this experiment and Corollary 4.2.2.1 are in agreement with conclusions drawn from the classic identifiability analysis technique of matching the transfer function of the model with an estimated one (see Section 4.4.2).

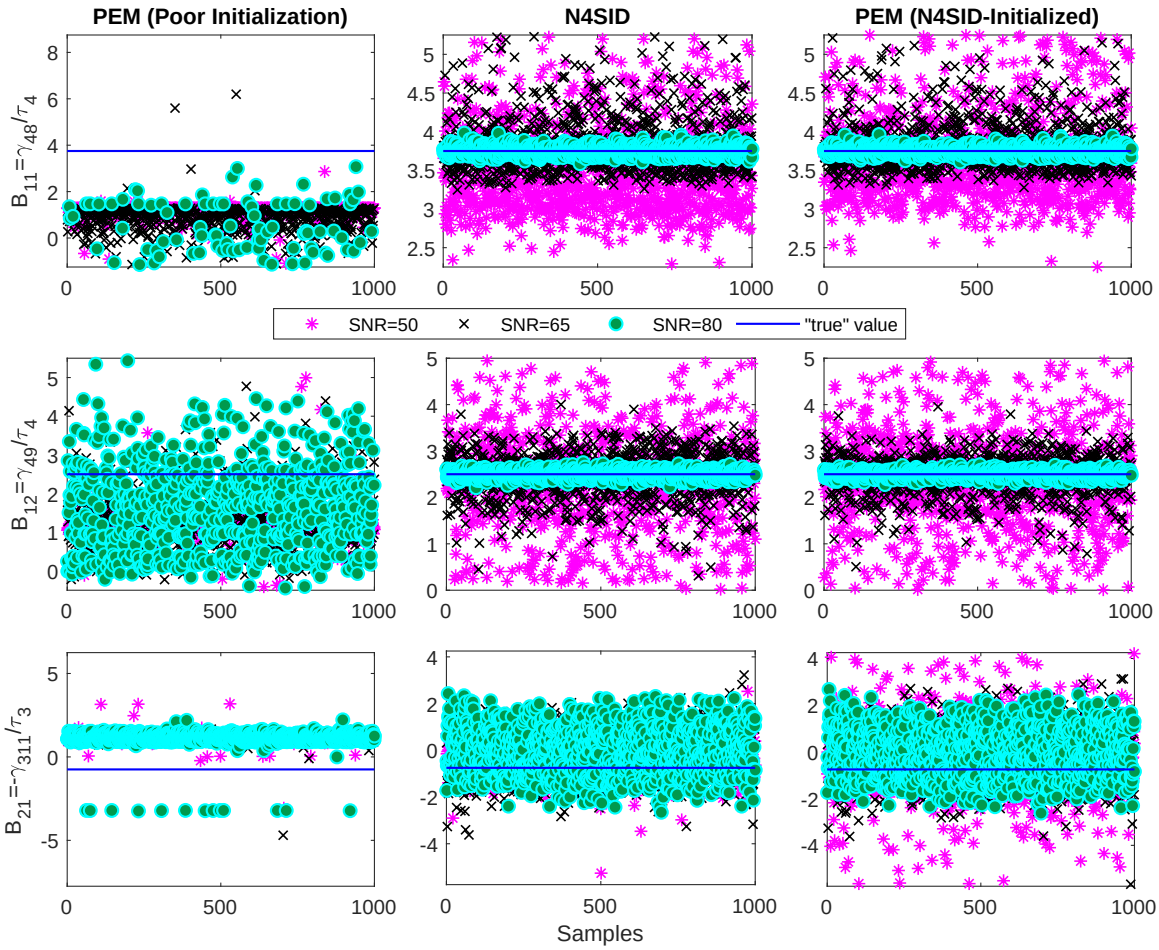

Figure 4.3: Visualization of the Partial Identifiability in $B(\theta_p)$ of the OCSE Model in (4.25) under Three Different Signal-to-noise Ratio (SNR) Scenarios Per Corollary 4.2.2.1. Solver Settings and Random Elements (T_f and output noise) Retained in the Estimated Hypothetical Models for Verified Reproducibility.

4.4 *Just Walk*: OCSE-SCT Semiphysical Identification Results

This section reports new results extending from the prior work in Freigoun *et al.* and dos Santos *et al.*. Input-output data are used to estimate individual OCSE-SCT models. Simulation response comparisons against used participants data (data retrieved from Freigoun *et al.*, 2017; Mercere, 2017 with high accuracy; see Appendix A) are presented in Figure 4.4. For each participant, three models are estimated to generate and compare predictions of individual behaviors [steps/day] across dif-

ferent datasets: First, the standard N4SID model is estimated. Second, using the YALMIP toolbox J. Löfberg, 2004, this first model is applied to (4.17) for estimating an eigenvalue-constrained model over the three common LMI regions in Miller and De Callafon, 2013 with α and β ensuring stability, damped response, and physical realization, i.e.,

$$\alpha = \begin{bmatrix} (1-\delta_s)I_2 & 0 & 0 \\ 0 & 2\delta_r I_2 & 0 \\ 0 & 0 & 2\delta_p I_2 \end{bmatrix}, \quad \beta = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.26}$$

Finally, with the absence of prior knowledge on parameter ranges, the constrained model and the featured SD formulation in Section 4.2 are used to initialize the standard PEM estimator to enforce the OCSE structure in Example 4.3.2. The NRMSE standard defined in (3.5) is used to quantify the goodness of fit.

4.4.1 Model Validation

Once a certain model is developed and estimated (particularly black- and grey-box models), a central step in system identification is *model validation*. Ideally, the experimental data are split into two segments (or concatenated into two larger segments) for model estimation and validation.

Typically, the ultimate deliverable in model validation is to show that, at some level, the model in question has not been falsified by experimental data (Ljung, 1999). When comparing models and their predictions against measured input-output data, from an *agreement* point of view, one may be curious to run a simple Bland-Altman (BA) analysis. In areas such as analytical chemistry, the BA plot is often used to examine the *agreement* between measures of the same construct using two distinct methods/devices (note the use of the term *agreement* and not *correlation*; see Jinyuan

et al., 2016). As a word of caution, we cite O'Connor *et al.*, 2011 in that "the Bland-Altman method should not be used in regression cross-validation studies." Nonetheless, since the BA plot is ultimately a difference plot (i.e., a scatter plot of model *error* in this context), the BA plot can give a crude assessment for the *bias* and *variance* across different models as they are compared against measured data. Figure 4.5 presents the BA plots for each of the established models in Figure 4.4 (participant B). It is noted that the main ensuing observations from BA *agreement* plots are reproducible from participant A.

In Figure 4.5, the four models of Participant B are considered (four columns): the Optimal ARX model, initial N4SID subspace model, the eigenvalue-constrained derivative of the first N4SID subspace model, and structured SCT-OCSE model. In the two rows, the corresponding difference and agreement BA plots are provided with a summary of numerical values. It is clearly shown that the initial subspace model is drastically improved following model reduction and the incorporation of structural modeling insights imported from Social Cognitive Theory using the SD-PEM algorithm. It is also evident the structured OCSE delivers the best *bias-variance* trade-off, and the highest agreement with measured data (i.e., steps/day).

In terms of using statistical *correlation* in control-oriented model validation, the conventional approach in system identification includes the use the cross-validation dataset for residual analysis. With an established degree of confidence, it can be determined if the concerned model is falsified by the data via examining auto-correlation of the residuals and their cross-correlation with measured input signals. In Figure 4.6, we rely on 99% confidence intervals (i.e., 2.576σ) for a residual analysis on the validation data to conclude that the identified OCSE models are not falsified by the data (as opposed to ARX and unconstrained N4SID models for participant B). In other words, we cannot say that the OCSE models "have not picked up all dynamics" from

$u(t)$ on $\eta_4(t)$ (note the double negation as pointed in Ljung, 1999).

Overall, using simulation plots, BA plots, and residual analysis in Figures 4.4, 4.5, and 4.6 (respectively) we note the sensible improvements when prior knowledge is incorporated into the model (e.g., when LMI constraints in (4.17) are applied to the N4SID subspace model) and, ultimately, best improvements when models are reduced to conform with the OCSE structure in Figure 4.2, estimated via SD-PEM.

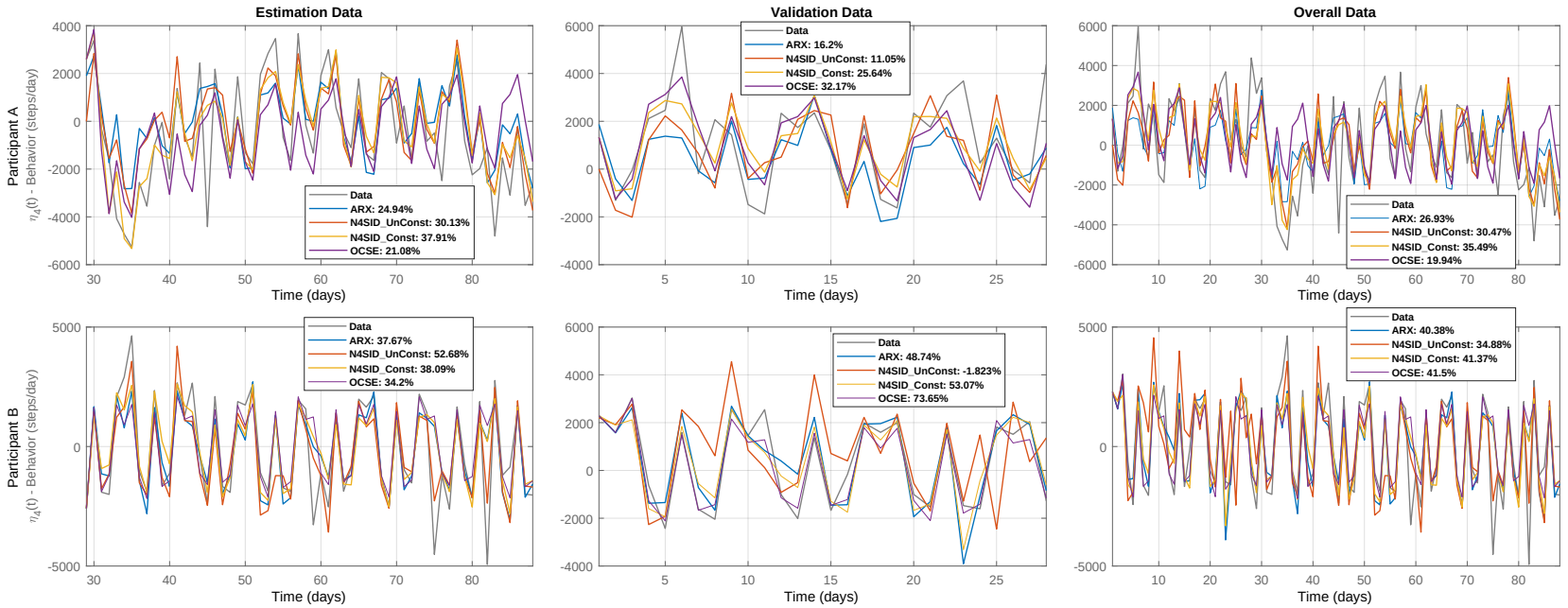

Figure 4.4: Simulations of the optimal ARX, Standard N4SID, Eigenvalue-constrained N4SID (Realizable) From (4.17) and (4.26), and SD-initialized OCSE-SCT Models for Two Representative *Just Walk* Participants In Freigoun *et al.*, 2017; Mercere, 2017 (See Appendix A). The Simulations Feature the Predicted Output $\eta_4(t)$ (*Behavior*) Against Input-Output Data in the Estimation, Validation, and Overall Data. Improving from Freigoun *et al.*, 2021b, Optimal ARX Models were Added and Other Models for Participant A were Re-estimated Following improved settings. *The Reader May Use the "zoom" Functionality to Enlarge This Page and View Values in the Electronic Version of This Dissertation.*

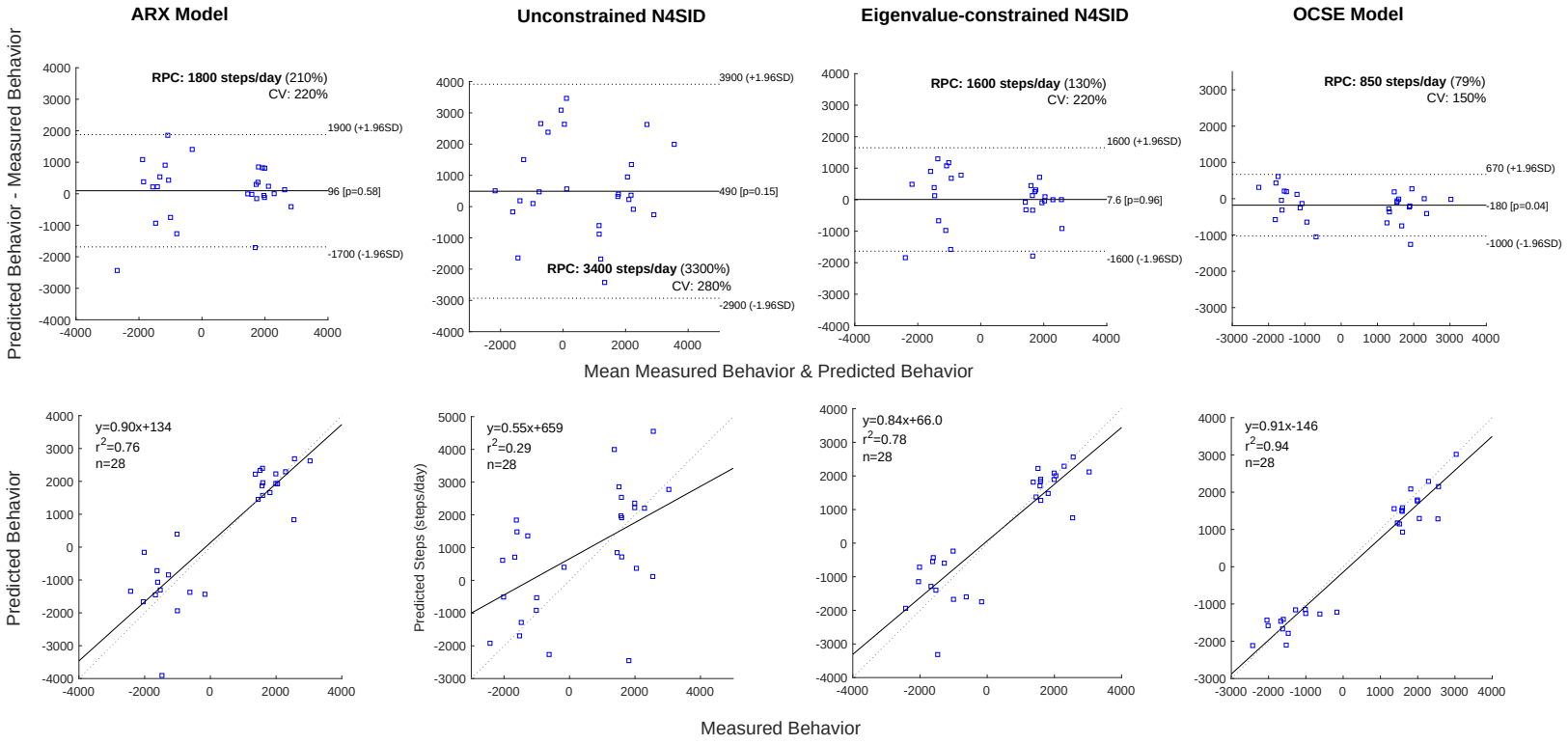

Figure 4.5: Bland-Altman *Agreement* Plots for the Four Estimated Behavioral Models of Participant B from *Just Walk* (Figure 4.4): Optimal ARX, Unconstrained N4SID, Eigenvalue-constrained N4SID, and the Semiphysical OCSE Model of Social Cognitive Theory. *Software*: Klein, 2014.

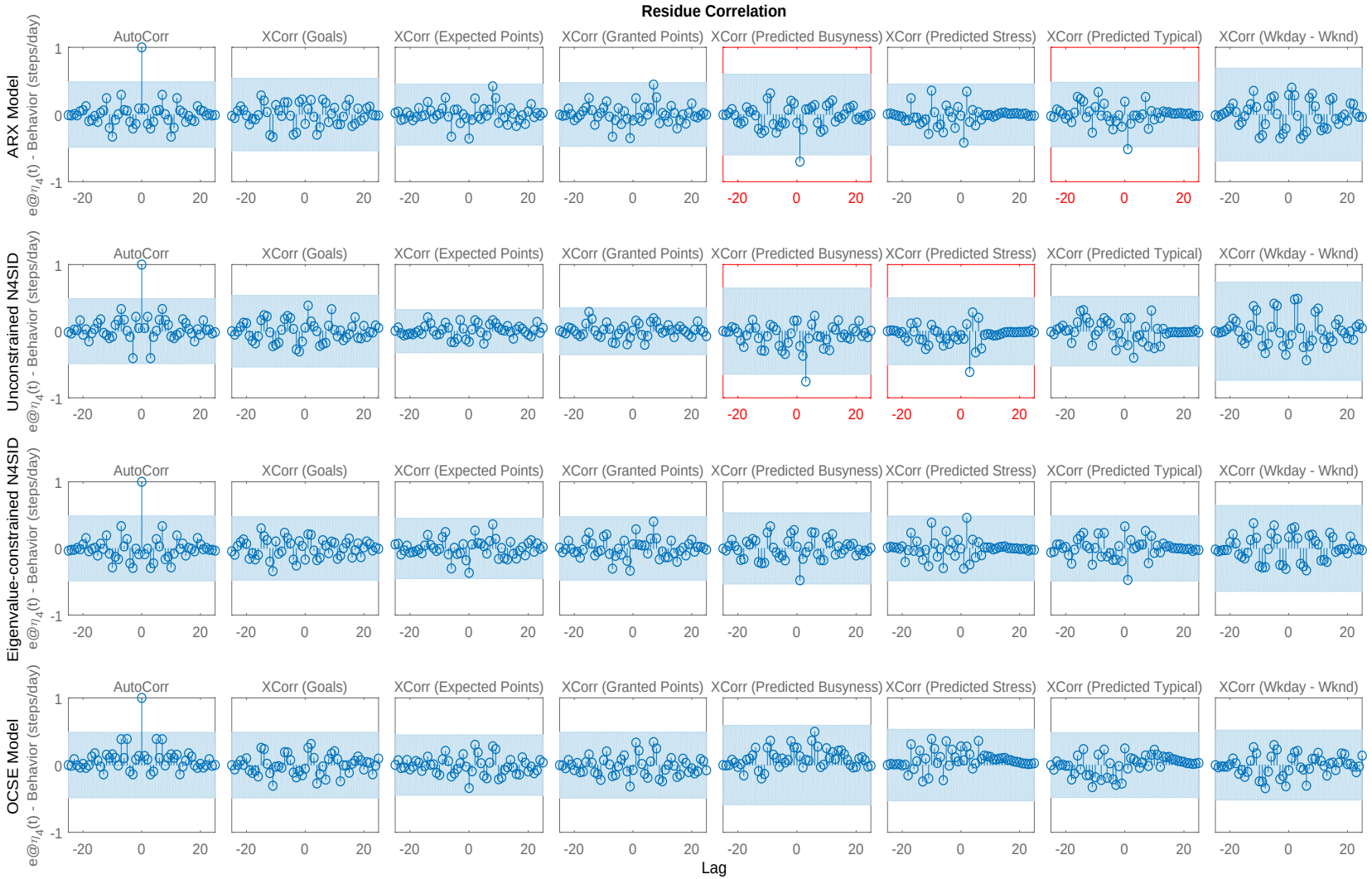

Figure 4.6: Residual Analysis Plots for Model Validation. Auto-correlation of Residuals and Cross-Correlations with Input Signals are Plotted Against 99% Confidence Intervals for the Four Estimated Behavioral Models of Participant B from *Just Walk* (Figures 4.4 and 4.5): Optimal ARX, Unconstrained N4SID, Eigenvalue-constrained N4SID, and the Semiphysical OCSE Model of Social Cognitive Theory. Frames in Red (Color) Emphasize Model 'Falsification' by Data.

4.4.2 OCSE Structure: Identifiability Analysis

In this section, an identifiability analysis on the OCSE model in Figure 4.2 is performed. First, we reduce the unknown state-space variables in (5.2b) as follows:

$$A = \begin{bmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & 0 \\ a_6 & 0 & a_7 \end{bmatrix}, \quad B = \begin{bmatrix} b_1 & b_2 & 0 & 0 & 0 & 0 & 0 & 0 \\ b_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & b_4 & b_5 & b_6 & b_7 & b_8 & b_9 \end{bmatrix}, \quad C = [1 \quad 0 \quad 0] \tag{4.27}$$

with the aim of establishing the degree(s) of freedom in the model. A classical technique is to obtain the transfer function matrix (input-output unique) in terms of the variables $\{a_i\}_{i=1:7}$ & $\{b_j\}_{j=1:9}$, and then verify if these can be retrieved uniquely from input-output data. Using

$$H(s) = C(sI - A)^{-1}B + D$$

determines the following transfer function matrix $H(s)$

$$H(s) = \left[\begin{array}{c} \frac{-b_1 s^2 + (a_5 b_1 + a_7 b_1 - a_2 b_3)\, s + (a_2 a_7 b_3 - a_5 a_7 b_1)}{s^3 - (a_1 + a_5 + a_7)\, s^2 + (a_1 a_5 + a_1 a_7 + a_5 a_7 - a_2 a_4 - a_3 a_6)\, s + (a_2 a_4 a_7 + a_3 a_5 a_6 - a_1 a_5 a_7)} \\ \frac{-b_2 s^2 + b_2\,(a_5 + a_7)\, s - (a_5 a_7 b_2)}{s^3 - (a_1 + a_5 + a_7)\, s^2 + (a_1 a_5 + a_1 a_7 + a_5 a_7 - a_2 a_4 - a_3 a_6)\, s + (a_2 a_4 a_7 + a_3 a_5 a_6 - a_1 a_5 a_7)} \\ \frac{-a_3 b_4 s + (a_3 a_5 b_4)}{s^3 - (a_1 + a_5 + a_7)\, s^2 + (a_1 a_5 + a_1 a_7 + a_5 a_7 - a_2 a_4 - a_3 a_6)\, s + (a_2 a_4 a_7 + a_3 a_5 a_6 - a_1 a_5 a_7)} \\ \frac{-a_3 b_5 s + (a_3 a_5 b_5)}{s^3 - (a_1 + a_5 + a_7)\, s^2 + (a_1 a_5 + a_1 a_7 + a_5 a_7 - a_2 a_4 - a_3 a_6)\, s + (a_2 a_4 a_7 + a_3 a_5 a_6 - a_1 a_5 a_7)} \\ \frac{-a_3 b_6 s + (a_3 a_5 b_6)}{s^3 - (a_1 + a_5 + a_7)\, s^2 + (a_1 a_5 + a_1 a_7 + a_5 a_7 - a_2 a_4 - a_3 a_6)\, s + (a_2 a_4 a_7 + a_3 a_5 a_6 - a_1 a_5 a_7)} \\ \frac{-a_3 b_7 s + (a_3 a_5 b_7)}{s^3 - (a_1 + a_5 + a_7)\, s^2 + (a_1 a_5 + a_1 a_7 + a_5 a_7 - a_2 a_4 - a_3 a_6)\, s + (a_2 a_4 a_7 + a_3 a_5 a_6 - a_1 a_5 a_7)} \\ \frac{-a_3 b_8 s + (a_3 a_5 b_8)}{s^3 - (a_1 + a_5 + a_7)\, s^2 + (a_1 a_5 + a_1 a_7 + a_5 a_7 - a_2 a_4 - a_3 a_6)\, s + (a_2 a_4 a_7 + a_3 a_5 a_6 - a_1 a_5 a_7)} \\ \frac{-a_3 b_9 s + (a_3 a_5 b_9)}{s^3 - (a_1 + a_5 + a_7)\, s^2 + (a_1 a_5 + a_1 a_7 + a_5 a_7 - a_2 a_4 - a_3 a_6)\, s + (a_2 a_4 a_7 + a_3 a_5 a_6 - a_1 a_5 a_7)} \end{array} \right]^T$$

and, consequently, the following set of equations follow:

$$b_1 = -\alpha_1 \tag{4.28}$$
$$(a_5 + a_7)b_1 - a_2 b_3 = \alpha_2 \tag{4.29}$$
$$a_2 a_7 b_3 - a_5 a_7 b_1 = \alpha_3 \tag{4.30}$$
$$b_2 = -\alpha_4 \tag{4.31}$$
$$(a_5 + a_7)b_2 = \alpha_5 \tag{4.32}$$
$$a_5 a_7 b_2 = -\alpha_6 \tag{4.33}$$

$$a_3 b_4 = -\alpha_7 \tag{4.34}$$
$$a_3 a_5 b_4 = \alpha_8 \tag{4.35}$$
$$a_3 b_5 = -\alpha_9 \tag{4.36}$$
$$a_3 a_5 b_5 = \alpha_{10} \tag{4.37}$$
$$a_3 b_6 = -\alpha_{11} \tag{4.38}$$
$$a_3 a_5 b_6 = \alpha_{12} \tag{4.39}$$
$$a_3 b_7 = -\alpha_{13} \tag{4.40}$$
$$a_3 a_5 b_7 = \alpha_{14} \tag{4.41}$$
$$a_3 b_8 = -\alpha_{15} \tag{4.42}$$
$$a_3 a_5 b_8 = \alpha_{16} \tag{4.43}$$
$$a_3 b_9 = -\alpha_{17} \tag{4.44}$$
$$a_3 a_5 b_9 = \alpha_{18} \tag{4.45}$$
$$a_1 + a_5 + a_7 = -\alpha_{19} \tag{4.46}$$
$$a_1 a_5 + a_1 a_7 + a_5 a_7 - a_2 a_4 - a_3 a_6 = \alpha_{20} \tag{4.47}$$
$$a_2 a_4 a_7 + a_3 a_5 a_6 - a_1 a_5 a_7 = \alpha_{21} \tag{4.48}$$

where α_i are the identified numerical coefficients of the estimated $H(s)$. It is highlighted that, in alignment with Corollary 4.2.2.1, $b_1 = B_{11} = \frac{\gamma_{48}}{\tau_4}$ and $b_2 = B_{12} = \frac{\gamma_{49}}{\tau_4}$ are indeed uniquely identifiable from input-output data per (4.28) and (4.31) (also see Figure 4.3). In addition, a moment's glance at equations (4.34)-(4.45) renders a_5 known, which when plugged into (4.32) given (4.31) also determines a_7. Further, with known a_5 and a_7, a_1 is immediately obtained from (4.46).

Given known $a_{(1,5,7)}$ and $b_{(1,2)}$, and unknowns $a_{(2,3,4,6)}$ and $b_{(3-9)}$, it can be noted that the pairs $\{a_2, b_3\}, \{a_3, b_{(4-9)}\}, \{a_2, a_4\}$, and $\{a_3, a_6\}$ are coupled in equations (4.29), (4.30), (4.34)-(4.45), (4.47), and (4.48). Hence, for a unique determination of the structured model which yields global identifiability, at least two model parameters must be known *a priori*: at least one parameter must belong in the set $\{a_2, a_4, b_3\}$, and at least one another must belong in the set $\{a_3, a_6, b_{(4-9)}\}$. The following are a few examples:

i. **Given a_2 and a_3:** given a_2, b_3 is obtained from (4.29). Also, given a_3, $b_{(4-9)}$ are obtained from (4.34)-(4.45). Finally, with a_2 and a_3, equations (4.47) and (4.48)

yield a_4 and a_6. Hence, the system becomes identifiable given (a_2, a_3).

ii. **Given a_4 and a_6:** Plugging known a_4 and a_6 into equations (4.47) and (4.48) determines a_2 and a_3; the rest follows similar to (i.).

iii. **Given b_3 and any of $b_{(4-9)}$:** Equation (4.29) is used to determine a_2, and any of (4.34)-(4.45) can be used to determine a_3; the rest follows similarly from (i.).

This analysis makes quite clear that there are two degrees of freedom in the OCSE model, hence the model structure is unidentifiable, which entails an infinite number of models with identical and different structures that may be observationally equivalent. It is also clear that (iii.) corroborates Corollary 4.2.2.1 in that when all model states are directly observed (measured), the semiphysical OCSE model structure is indeed globally identifiable.

4.5 Chapter Summary

In this chapter, it was shown how a spectral decomposition of identified black-box models enables the formulation of constrained eigenvalue problems for estimating grey-box identification problems of a linear structure. The partial identifiability of a specific class of structures was discussed and a visualization from a numerical experiment was provided. Moreover, a further simplified OCSE model of Social Cognitive Theory was introduced as an extension from prior work in Chapter 3. This OCSE model was estimated and validated using input-output data. As is the case with existing literature, it may follow good practice if an explicit constraint is considered in (4.22) for managing the condition number of T_f during the search. While the proposed J_{SD} is still non-convex, local solutions with well-conditioned T_f matrices still

provide sub-optimal matches for the underlying system characteristics and thus can produce good candidate models for initializing PEM. With randomly-initialized J_{SD}, it is crucial to evaluate results from multiple runs, or use the multistart functionality of existing global optimization software. Finally, it is noted that the presented method can be extended to include and handle *quadratic* structures; the reader is referred to Section 6.3.4 for more insights.

Chapter 5

HYBRID MODEL PREDICTIVE CONTROL FOR CLOSED-LOOP INTERVENTION SIMULATIONS OF PARTICIPANT-VALIDATED SOCIAL COGNITIVE THEORY MODELS

5.1 Background

There is strong evidence highlighting the association between sedentary lifestyle and the increased risk for type 2 diabetes, cardiovascular disease, and cardiovascular and all-cause mortality (Wilmot *et al.*, 2012). Furthermore, McTiernan, 2008 not only attributed %25 of cancer cases worldwide to obesity and sedentarism, but also pointed out findings in several types of cancer indicating improved prognoses among individuals diagnosed with cancer and remained physically active. Motivated by the increasing access to affordable high-precision pedometers (i.e., motion sensors), the goal of maintaining $10,000$ steps/day has gained more popularity and was recommended for "apparently" healthy adults (Tudor-Locke and Bassett, 2004). A *sedentary* lifestyle index is suggested as $< 5,000$ steps/day, whereas activity in the range of $5,000$-$7,499$ steps/day is considered *low-active*, $7,500$-$9,999$ is classified as *somewhat active*, and $10,000$-$12,499$ steps/day is *active*. The aforementioned serious health ramifications of the sedentary way of life, together with the abundance of advanced mHealth technologies, have created an impetus to the development of improved interventions that are individually-tailored, adaptive, scalable, and cost-effective.

Just Walk, an innovative single-subject intervention experiment (open-loop) designed based on system identification principles, provided a unique opportunity for the

estimation and validation of individual dynamical systems (grey-box) models of Social Cognitive Theory (SCT) using input-output participant data (Hekler *et al.*, 2018; Freigoun *et al.*, 2017); see Chapters 3 and 4 for more elaboration on *Just Walk*.

Further, featured closed-loop physical activity intervention simulations in this work rely on a Hybrid Model Predictive Control (HMPC) controller design strategy. The general HMPC formulation in Bemporad and Morari, 1999; Bemporad, 2004 and hybrid decision rules in Martín *et al.*, 2016a are considered in the scope of this work. Similar to behavior "initiation" and "maintenance" phases in Martín *et al.*, 2016a, a controller reconfiguration strategy is proposed for the design of *multi-phase* interventions that optimally drive positive behavior change *gradually* over time. Generated closed-loop performance simulations in this work provide a proof of concept for the amenability of established methods in system identification and control to behavior-change problems.

The rest of this chapter is organized as follows: Section 5.2 features a reduced dynamical systems model of Social Cognitive Theory and reviews relevant prior system identification work. Section 5.3 reviews an existing HMPC formulation used for closed-loop intervention design. Finally, closed-loop performance simulations of a participant-validated model under different scenarios and controller configurations are provided in Section 5.4.

5.2 Grey-box Identification of *Just Walk*

In this section, a brief background of prior control-relevant identification work leading to participant-validated behavioral models is presented. With the development of a dynamic process model being an inherent requirement for MPC design, a reduced behavioral model lending its theoretical basis from Social Cognitive Theory is considered. The estimation and validation of this model from retrieved previously

published *Just Walk* input-output data is featured (see Appendix A).

5.2.1 Behavioral Process Model

In an attempt to translate a narrated presentation of SCT into a simple state-space realization amenable to theory-testing and capable of predicting behavior to a practically useful extent, Martín *et al.*, 2014 developed a sixth-order grey-box (i.e., semiphysical) SCT model using fluid analogy to capture reciprocal determinism and other key elements of the theory. A simplified version of this model (fourth-order) was published in Martín *et al.*, 2015; another fifth-order variant was featured in Freigoun *et al.*, 2017 and dos Santos *et al.*, 2018. Later, in Freigoun *et al.*, 2021b, a number of practical challenges with data and other factors imposed by the design and subsequent analyses of the pilot *Just Walk* study propelled the need for a more 'pragmatic' and parsimonious model by applying further simplifications leading to the 'core,' Operant Conditioning–Self-efficacy (OCSE) model of Social Cognitive Theory illustrated in Fig. 4.2. To name a few, these challenges and considerations stem from measurement design, data length, frequency content of relevant signals, estimated model order by inspection of singular values of a projection matrix computed intermediately in subspace analysis (suggesting third-order dynamics), optimal black-box model structures (also favoring third-order models; see Freigoun *et al.*, 2017 & dos Santos *et al.*, 2018), and the lack of sufficient prior knowledge with respect to parameter values in the original model. While most of these challenges can be alleviated through improved future *Just Walk*-like experiments, we meanwhile consider the third-order OCSE model in the scope of this chapter for the purpose of HMPC-based closed-loop intervention design. Model order determination using subspace analysis (e.g., N4SID) can be found in Theorems 2 and 8 of Van Overschee and De Moor, 2012. Further, featured closed-loop physical activity intervention simulations in this work rely on a

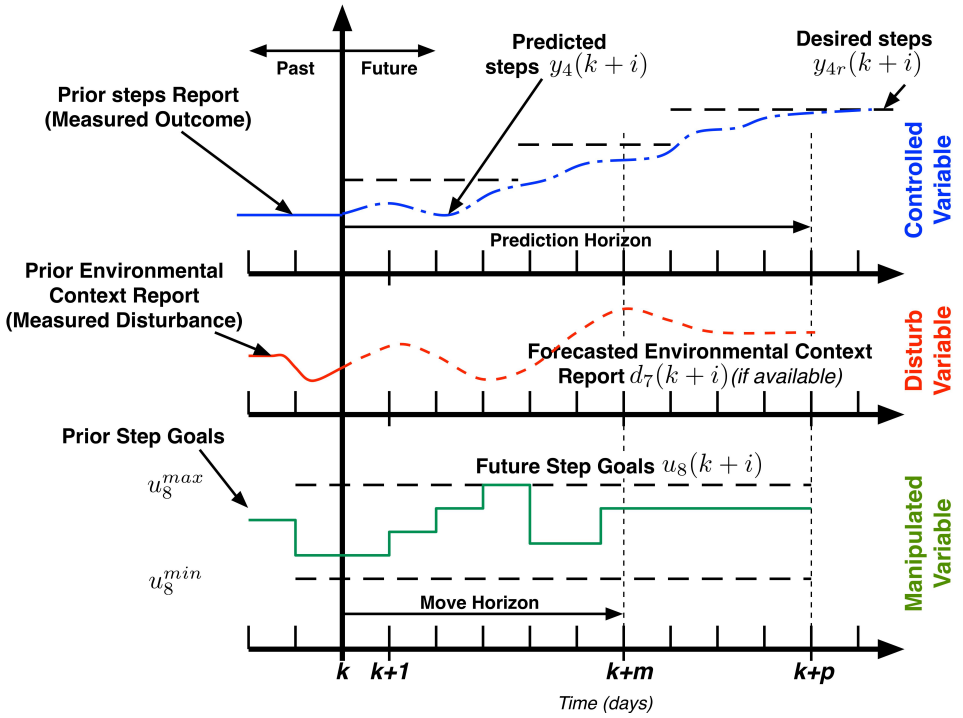

Figure 5.1: Conceptual Application of the Receding Horizon Control Strategy to the Physical Activity Behavioral Problem (Hekler *et al.*, 2018).

Hybrid Model Predictive Control (HMPC) controller design. A conceptualization of the HMPC, receding horizon strategy applied to physical activity interventions is illustrated in Figure 5.1. Similar to behavior "initiation" and "maintenance" phases in Martín *et al.*, 2016a, a controller reconfiguration strategy is proposed for the design of *multi-phase* interventions that optimally drive positive behavior change *gradually* over time. Generated closed-loop performance simulations in this work provide a proof of concept for the amenability of behavior-change problems to established methods in system identification and control.

OCSE Model

The semiphysical OCSE model in Figure 4.2 features a two-loop model consisting of *operant conditioning* (i.e., positive reinforcement/"negative punishment" learning) and *self-efficacy* loops. The reader is referred to Martín *et al.*, 2014 and Riley

et al., 2015 for more elaborate definitions of SCT constructs and signals relevant to this model. Applying the conservation of mass principle to the fluid flow model in Figure 4.2 yields the following set of first-order differential equations:

$$\tau_4 \frac{d\eta_4(t)}{dt} = \gamma_{48}\xi_8(t) + \gamma_{49}\xi_9(t) + \beta_{43}\eta_3(t) + \beta_{45}\eta_5(t) - \eta_4(t) \tag{5.1a}$$

$$\tau_3 \frac{d\eta_3(t)}{dt} = \gamma_{311}\xi_{11}(t) + \beta_{34}\eta_4(t) - \eta_3(t) \tag{5.1b}$$

$$\tau_5 \frac{d\eta_5(t)}{dt} = \gamma_{510}\xi_{10}(t) + \sum_{j=1}^{n_d} \gamma_{7_j}\xi_{7_j}(t) + \beta_{54}\eta_4(t) - \eta_5(t) \tag{5.1c}$$

which can be represented by the following state-space equations

$$\dot{x}(t) = Ax(t) + Bu(t) \tag{5.2a}$$

$$y(t) = Cx(t) \tag{5.2b}$$

with system matrices defined as in (4.25).

5.2.2 Just Walk: Input Signal Design

As a part of the *Just Walk* open-loop intervention design, orthogonal-in-frequency multisine excitations of independent intervention dosages (i.e., $u_n(t) = \xi_n(t);\ n = \{8, 9\}$) were generated for each individual subject based on their baseline physical activity. For more details on this "zippered" spectra design, the reader may review Section 3.3 for further elaboration, as well as Martín *et al.*, 2015 in which an input signal design procedure motivated by "patient-friendly" constraints is provided. The input signal *granted points*, $u_{10}(t) = \xi_{10}(t)$, is constructed by using an "If-Then" rule defined by the function

$$u_{10}(k) \triangleq \begin{cases} u_9(k-1) & y(k) \geq u_8(k-1) \\ 0 & y(k) < u_8(k-1) \end{cases} \tag{5.3}$$

where $k \in \mathbb{N}$ is the discrete-time sampling instance.

5.2.3 Model Estimation and Validation

***Just Walk* Identifiability Limitations**

In experimental settings where the structure of the model itself is of a particular significance, or when unknown model parameters carry some physical significance or certain implications about the underlying phenomenon, one potential requirement is the acquisition of informative measurements of *all* states of the minimal model representation. This requirement enables the estimation of a globally-optimal, unique model as a result of the guaranteed global identifiability. Freigoun *et al.*, 2021b presented a subspace-based spectral decomposition algorithm for estimating structured continuous-time state-space models conforming with (5.2). Maintaining two common practical assumptions, the proposed method can be of a particular value for data-driven theory testing experiments since the convexity of the estimation problem is extended from obtaining a fully-parameterized subspace model to a structured continuous-time one such as the OCSE model in Figure 4.2. Thus, in the absence of prior knowledge concerning relevant parameter values, an important experiment design requirement in future *Just Walk*-like experiments is to reliably measure responses of all SCT states in the considered SCT model. Hence, for a more complete validation of the OCSE model structure itself, measurements capturing *self-efficacy* $\eta_3(t)$ and *behavioral outcomes* $\eta_5(t)$ are as essential as the measurable *behavior* $\eta_4(t)$, i.e., $C = I$.

Selected Participant Results

Despite being unidentifiable in the context of *Just Walk*, individual OCSE models are estimated and validated using an SD-PEM method in Freigoun *et al.*, 2021b and digitized published *Just Walk* participant data in Freigoun *et al.*, 2017 (see Appendix

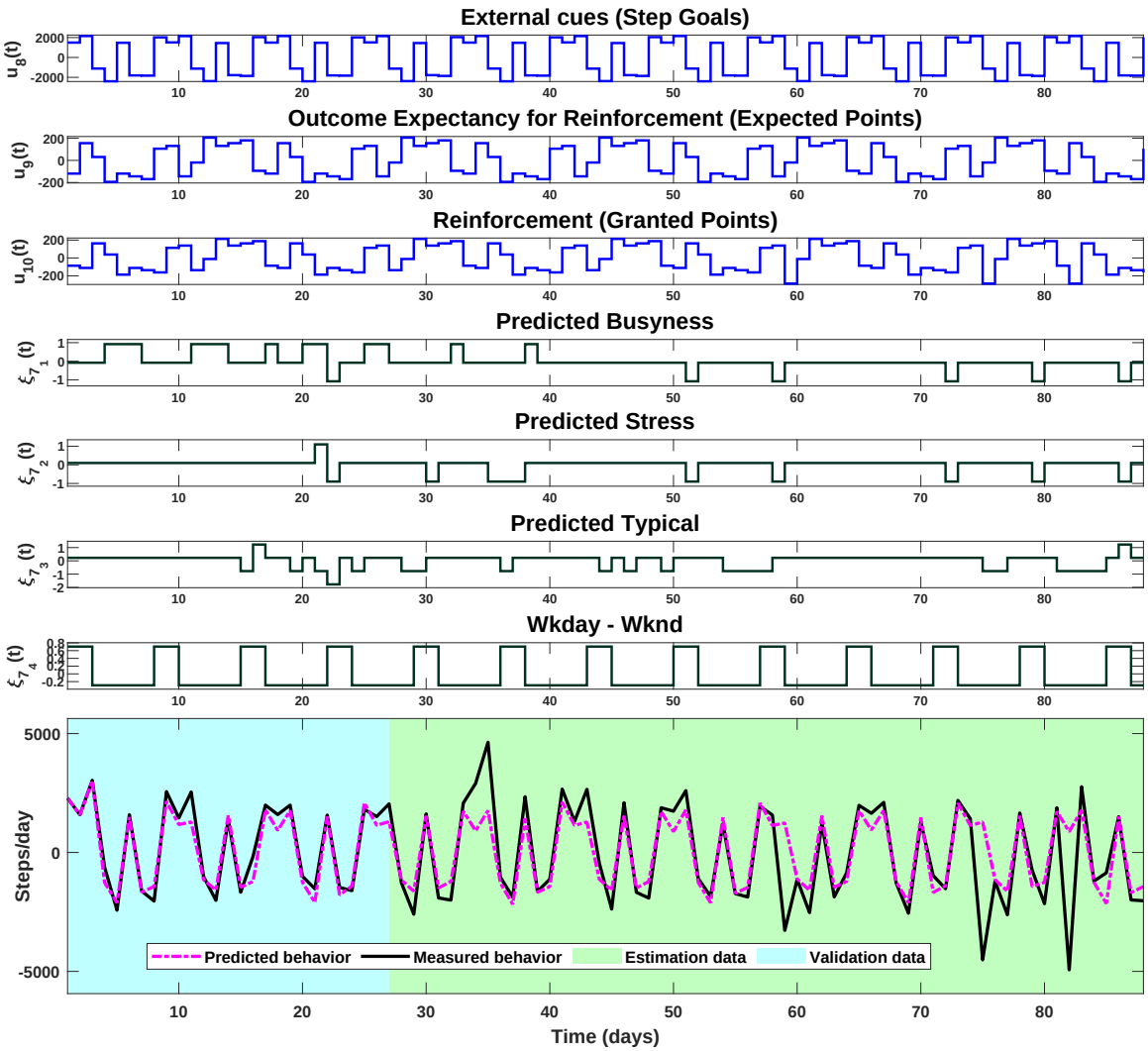

Figure 5.2: Time Series Plot Illustrating Input Signals, Measured Changes in Behavior (Output) [Steps/Day], and the Predicted Behavior from the OCSE Model Using Retrieved Input-Output Participant Data in Freigoun *et al.*, 2017 (see Appendix A) with $n_d = 4$. NRMSE Fits: Estimation (34.2%, Green Region); Validation (73.65%, Cyan Region); Overall Data (41.5%).

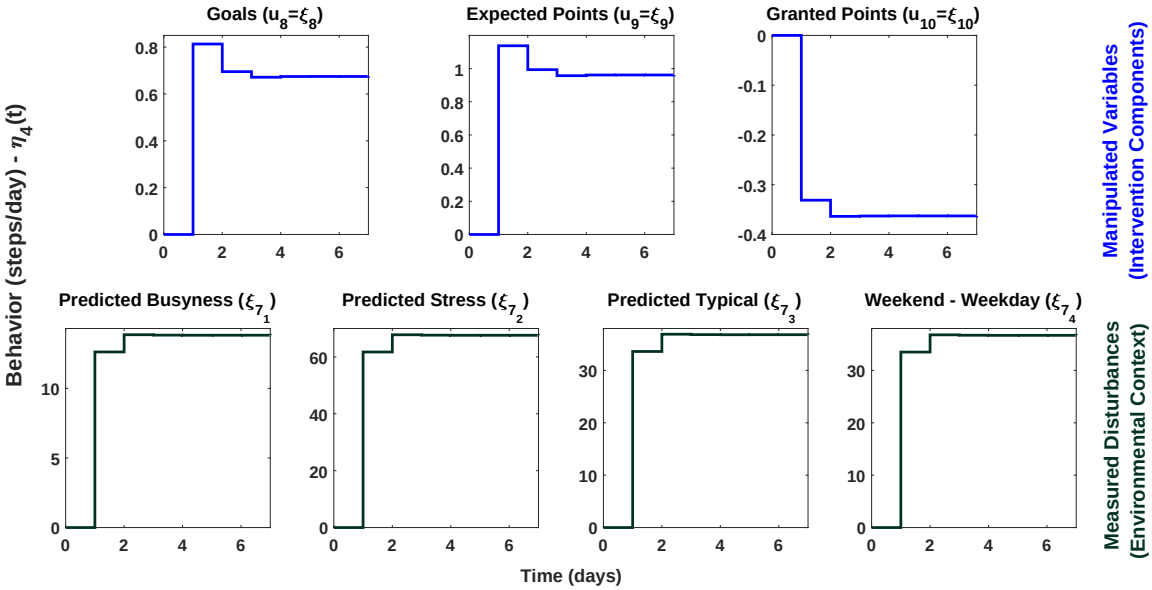

Figure 5.3: Step Responses of the Semiphysical OCSE Model for a *Just Walk* Participant over a 7-day (Week) Prediction Horizon.

A for accuracy). A simulation of a selected *Just Walk* participant model (OCSE structure) is depicted in Figure 5.2. As indicated in Figure 5.2, this participant's behavior is best predicted with the following disturbance signals: *predicted busyness*, *predicted stress*, *predicted typical*, and *weekday-weekend*. The first 27 input-output samples were reserved for model validation, and the following 61 samples were used for model estimation. An overall model NRMSE fit of 41.5% is reported in Freigoun *et al.*, 2021b, which highlights that the identified OCSE models using SD-PEM outperformed classical subspace models (i.e., standard and eigenvalue-constrained N4SID models).

Step Responses

While goodness of fit over validation data and residual analysis provide important criteria for model validation, step responses that are in keeping with common intuition are desirable. For the identified, participant-validated OCSE model featured

in Figures 4.2 and 5.2, step responses for manipulated inputs are shown in the top row of Figure 5.3, followed by responses to unit step changes in disturbance signals in the bottom row. Understandably, the *Just Walk* experiment length and input signals power and resolution, constrained by "patient-friendly" design considerations, can negatively impact confidence intervals. As a result, Section 5.4 is restricted to considering lower prediction and move horizons in the HMPC loop design; this is consistent with the short horizon controller tuning in Martín *et al.*, 2016a. Furthermore, as an additional value drawn from inspecting step responses, the likely directions and initial amplitudes depicted in Fig. 5.3 can guide the design of disturbance signals for the purposes of closed-loop performance evaluation (not in this chapter).

5.3 HMPC Framework

In this section, the main HMPC controller formulation adopted in this work is reviewed and relevance to the particular physical activity application setting is emphasized. With an HMPC design strategy with an embedded OCSE behavioral model, *setpoint tracking* (i.e., reaching and maintaining desired physical activity levels) is achieved by directly manipulating issued goals $u_8(t)$, expected points $u_9(t)$ for positive reinforcement, and indirectly via granted points per (5.3). *Measured disturbance rejection* is similarly designed for mitigating the effects from measured environmental context signals such as measured disturbances $\xi_{7_i}(t)$ featured in Figs. 5.2 and 5.3; correction for unmeasured disturbances and/or unmodeled dynamics is accomplished through state feedback. For the latter, Martín *et al.*, 2016a gives the example of "sickness of a family member" for a potentially immeasurable isolated (i.e., discrete) event. An additional real-life example from the *Just Walk* experiment is that one participant had reported migraine, accounting for one of the two sharpest drops in steps towards the end of the open-loop intervention as depicted in Fig. 5.2.

5.3.1 *Main HMPC formulation.*

In this section, we present a brief recapitulation of the optimal control problem formulation for hybrid systems in Bemporad and Morari, 1999 that describes the mechanism for the used `hybcon` controller object of the Hybrid Toolbox (Bemporad, 2004) in MATLAB® is presented. Moreover, operational constraints developed in Martín *et al.*, 2016a are also incorporated in the implementation of this particular application setting.

A hybrid linear system with n_u inputs (discrete and continuous), real and integer states, and n_y outputs subject to logical/discrete decisions can be described with the following Mixed Logical and Dynamical (MLD) representation (Bemporad and Morari, 1999):

$$x_{k+1} = Ax_k + B_1u_k + B_2\delta_k + B_3z_k + B_dd_k \tag{5.4a}$$

$$y_k = Cx_k + D_1u_k + D_2\delta_k + D_3z_k + d'_k \tag{5.4b}$$

$$E_2\delta_k + E_3z_k \leq E_1u_k + E_4x_k + E_5 \tag{5.4c}$$

where $x = [x_c^T \; x_\ell^T]^T$ is the hybrid system state vector, $x_c \in \mathbb{R}^{n_x^c}$ and $x_\ell \in \{0,1\}^{n_x^\ell}$; $u = [u_c^T \; u_\ell^T]^T$ is the system input with continuous and discrete/logical elements $u_c \in \mathbb{R}^{n_u^c}$ and $u_\ell \in \{0,1\}^{n_u^\ell}$; $y \in \mathbb{R}^{n_y}$ is the system output; d and d' are measured and unmeasured disturbances, respectively, in $\mathbb{R}^*$. $\delta \in \{0,1\}^{n_\delta}$ and $z \in \mathbb{R}^{n_z}$ are discrete/logical and continuous auxiliary variables that establishes the character of the concerned hybrid system.

With a hybrid system expressed in MLD structure, similar to classical MPC, the used hybrid optimal control problem in Bemporad, 2004 is formulated to optimize the sequence of control actions $\{u, \delta, z\}_0^N$ over an interval N. The given control law

minimizes the p-norm cost function J, i.e.,

$$\min_{\{u,\delta,z\}_0^N} \; J\left(\{u,\delta,z\}_0^N, x(t)\right) \tag{5.5}$$

where

$$J \triangleq \|Q_{xN}\left(x(N|t) - x_r\right)\|_p + \sum_{k=1}^{N} \|Q_x\left(x(k) - x_r\right)\|_p + \sum_{k=1}^{N} \|Q_u\left(u(k) - u_r\right)\|_p$$
$$+ \|Q_z\left(z(k|t) - z_r\right)\|_p + \|Q_y\left(y(k|t) - y_r\right)\|_p$$

subject to the mixed integer constraints established in (5.4c) and the following hard bounds:

$$x_{\min} \leq x_{k+i|k} \leq x_{\max}, \qquad i \in [1,\ N] \subset \mathbb{N} \tag{5.6a}$$

$$u_{\min} \leq u_{k+i} \leq u_{\max}, \qquad i \in [0,\ N-1] \subset \mathbb{I} \tag{5.6b}$$

$$y_{\min} \leq y_{k+i} \leq y_{\max}, \qquad i \in [0,\ N-1] \subset \mathbb{I} \tag{5.6c}$$

In (5.5), $*_r$ denote the reference signals/levels for states, inputs, continuous auxiliary variables, and outputs, respectively; $Q_* \triangleq Q_*^T \succeq 0$ are penalty weight matrices on the feedback error, energy/control signal, auxiliary continuous variables, and outputs, respectively. The reader is referred to Bemporad and Morari, 1999; Bemporad, 2004 for more elaboration and detailed notation of the HMPC problem in (5.5).

Similar to Nandola and Rivera, 2013, for *setpoint tracking* and *measured disturbance rejection*, the implemented formulation utilizes an adjustable Type-I discrete-time filter structure from Morari and Zafiriou, 1989, i.e.,

$$f(q, \alpha_{[r,d]}) = \frac{(1 - \alpha_{[r,d]})q}{q - \alpha_{[r,d]}} \tag{5.7}$$

where q is the forward-shift operator and $\alpha_{[r,d]} \in [0, 1)$ are adjustable tuning parameters governing the speed of response; a lower $\alpha_{[r,d]}$ drives a faster response and vice versa. While some measured disturbances of the OCSE model may be forecasted

(e.g., *weekday-weekend* signal $\xi_{7_4}(t)$), we assume for simplicity that all d_k signals are not forecasted in featured closed-loop performance simulations of Section 5.4.

Finally, it is noted that in the absence of a move suppression term in (5.5) that penalizes move sizes (useful in avoiding rapid change of intervention dosages), it is possible to mimic that effect in the Hybrid Toolbox via online programmatic controller reconfiguration at each time step. This may be achieved with an adjustment of hard constraints in (5.6b) or by considering soft constraints (i.e., ρ parameter of the Hybrid Toolbox, `Q.rho`).

5.3.2 HMPC-OCSE design considerations.

Particular to the design of the hybrid closed-loop system at hand, i.e., assignment of 'optimal' dosages of daily step goals and reward points, Martín *et al.*, 2016a established a set of operational constraints for predefined finite sets that include available intervention dosages; constraints for the realization of *granted points* per (5.3) were also given. To review, for the featured performance simulations in Section 5.4, the following sets of possible/feasible intervention dosages are defined

$$u_8(k) \in \mathcal{U}_8 \stackrel{\text{def}}{=} \left\{ c_8\nu_1, \ldots, c_8\nu_{n_{u_8}} \right\}, \tag{5.8a}$$

$$u_9(k) \in \mathcal{U}_9 \stackrel{\text{def}}{=} \left\{ c_9\nu_{n_{u_8}+1}, \ldots, c_9\nu_{n_{u_8}+n_{u_9}} \right\} \tag{5.8b}$$

where c_8 and c_9 are some established constants (e.g., c_8 is baseline behavior in (5.8a), c_9 is 100 points in (5.8b)) that are scaled up (or down) by value options ν_i. These constraints can be incorporated into the HMPC formulation in (5.4c) by introducing the logical and continuous auxiliary variables $\delta_j(k)$ and $z_j(k)$, respectively, such that

$$z_j(k) = c_{[8,9]}\nu_j\delta_j(k) \tag{5.9}$$

and

$$\sum_{j=1}^{n_{u_8}} \delta_j(k) = 1, \quad u_8(k) = \sum_{j=1}^{n_{u_8}} z_j(k) \tag{5.10}$$

$$\sum_{j=n_{u_8}+1}^{n_{u_8}+n_{u_9}} \delta_j(k) = 1, \quad u_9(k) = \sum_{j=n_{u_8}+1}^{n_{u_8}+n_{u_9}} z_j(k) \tag{5.11}$$

Furthermore, in Martín *et al.*, 2016a, additional logical and continuous auxiliary variables $\delta_{10}(k)$ and $z_{10}(k) = u_{10}(k)$ are introduced to enforce the "If-Then" definition of the *granted points* input $u_{10}(k)$ per (5.3). 'Big-M' formulations in Bemporad and Morari, 1999 are used, i.e., the high-level descriptions

$$[f_1(x) \stackrel{\text{def}}{=} y(k) - u_8(k-1) \leq 0] \rightarrow [\hat{\delta}_{10} = 1], \qquad \hat{\delta}_{10}(k) \stackrel{\text{def}}{=} 1 - \delta_{10}(k),$$

$$\{[\delta_{10}(k) = 1] \rightarrow [z_{10}(k) = u_9(k-1)]\} \wedge \{[\delta_{10}(k) = 0] \rightarrow [z_{10}(k) = 0]\}$$

establish the following set of constraints:

$$y(k) - u_8(k-1) \leq \delta_{10}(k)[y^{\max} - u_8^{\min}] \tag{5.12}$$

$$y(k) - u_8(k-1) \geq [1 - \delta_{10}(k)][y^{\min} - u_8^{\max}] \tag{5.13}$$

$$u_9(k-1) - z_{10}(k) \leq [1 - \delta_{10}(k)][u_9^{\max} - u_{10}^{\min}] \tag{5.14}$$

$$u_9(k-1) - z_{10}(k) \geq [1 - \delta_{10}(k)][u_9^{\min} - u_{10}^{\max}] \tag{5.15}$$

and

$$z_{10}(k) \leq \delta_{10}(k) u_{10}^{\max}, \quad z_{10}(k) \geq \delta_{10}(k) u_{10}^{\min} \tag{5.16}$$

Consistent with the notation in Bemporad and Morari, 1999, constraints (5.14) and (5.15) are produced with $f_2(x) \stackrel{\text{def}}{=} u_9(k-1) - u_{10}(k)$, $M_2 = u_9^{\max} - u_{10}^{\min}$, and $m_2 = u_9^{\min} - u_{10}^{\max}$; i.e., $[\delta_{10}(k) = 1] \rightarrow [y_2 \stackrel{\text{def}}{=} \delta_{10}(k) f_2(x) \stackrel{\Delta}{=} 0]$. Constraints (5.16) are directly constructed to reflect the logic in $[\delta_{10}(k) = 0] \rightarrow [z_{10}(k) = 0]$. For simulations presented in Section 5.4, the HYSDEL software tool (Torrisi and Bemporad, 2004) was used to conveniently build (5.4c) via the construction of (5.9)-(5.16) from the established high level descriptions of the hybrid OCSE closed-loop system.

5.4 Closed-loop HMPC Simulations

In this section we present HMPC closed-loop performance simulations for single- and multi-phase intervention strategies using the participant-validated OCSE model from *Just Walk* (featured in Fig. 5.2). Results following different HMPC configurations and tuning are illustrated and discussed. For simulations performed in this section, the response of the 'true' plant (human subject) is set to match that of the model's noise-corrupted output. In both cases (i.e., single- and multi-phase), unmeasured disturbances are Gaussian processes with $d'_k \sim \mathcal{N}(0, 400)$. Tuning parameters in setpoint and disturbance rejection (measured and unmeasured) are $\alpha_r = 0.95$, $\alpha_d = [0.1]_{n_d}$. In all presented simulations, measured disturbance signals (i.e., *environmental context* signals in Fig. 4.2) are set to equal the values from the identified *Just Walk* participant featured in Fig. 5.2 in both intervention scenarios. These signals are extended (with mirroring) to proceed beyond the *Just Walk* experiment duration with as much resemblance to reality as possible.

5.4.1 Single-phase intervention.

Fig. 5.4 features a single-phase, 90-day simulation in which the intervention path is direct from *sedentary* to *active* with no intermediate transitions. Responses of three w_{u_9} tunings of Q_u are shown; the higher w_{u_9}, the more suppression on issued expected points (and, consequently, total granted points) is applied throughout the intervention. When the system of points-to-money conversion in *Just Walk* is used (i.e., 500 points $\equiv$ \$1; Hekler *et al.*, 2018), Fig. 5.4 shows estimates for the granted/expected dollar amounts corresponding to each suppression setting of w_{u_9}. For example, when $w_{u_9} = 0.25$, the potential *expected* reward of the corresponding intervention is \$18, of which the individual is *actually* granted \$2.2 by 'cashing' the monetary equivalent

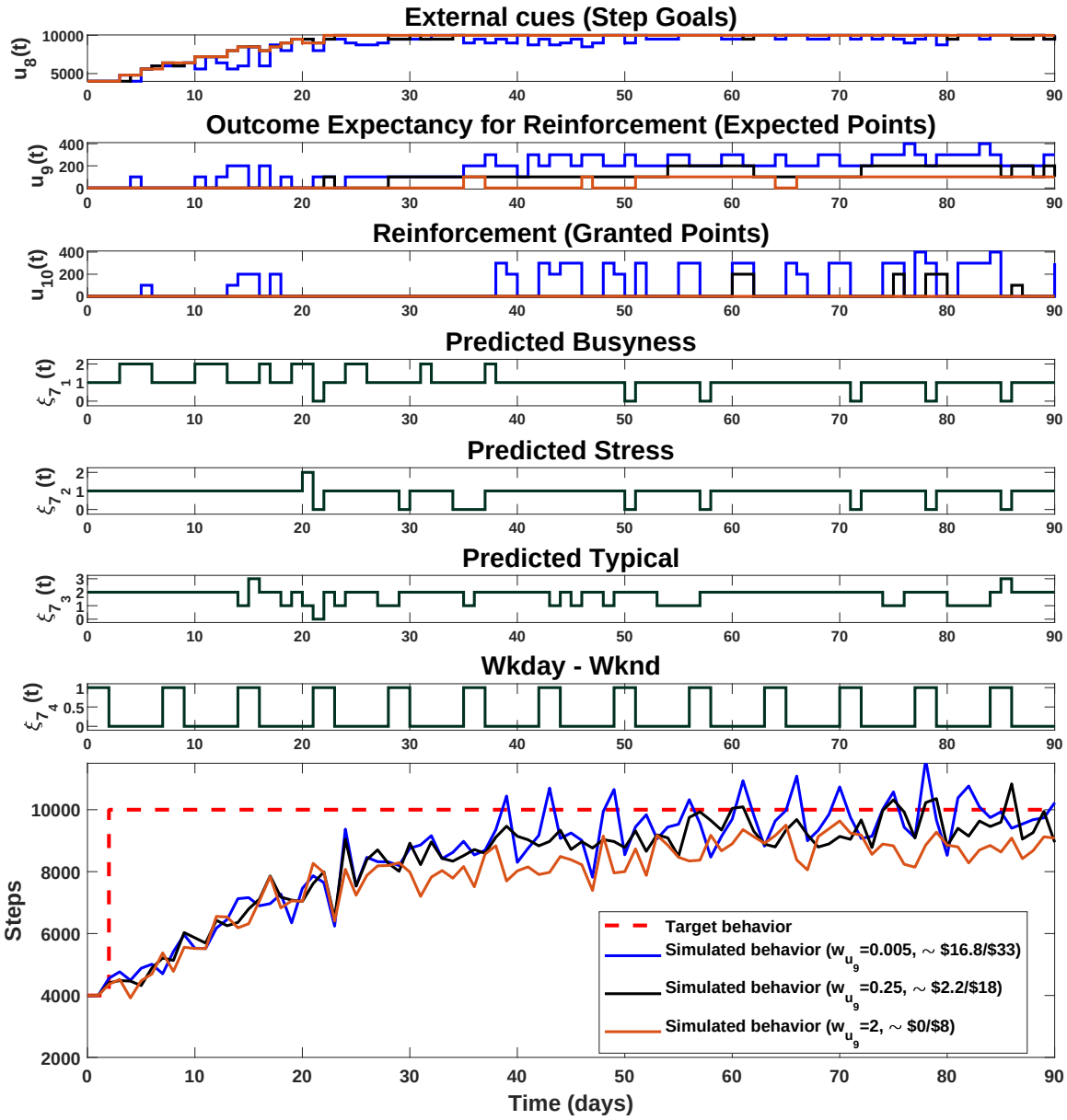

Figure 5.4: HMPC Closed-loop Performance Simulations (Single-phase Intervention) of a Participant-validated OCSE-SCT Model. HMPC Settings: Sampling Time $T_s = 1$ day, $N = 3$, $u_{\min} = [4000 \quad 0 \quad 0]^T$, $u_{\max} = [10000 \quad 500 \quad 500]^T$, $y_{\min} = 0$, $Q_y = 1$, $Q_u = \text{diag}\{0, [0.005, 0.25, 2], 0\}$, $c_8 = 4000$, $\nu_j|_{\mathcal{U}_8} = \{1, 1\frac{1}{8}, 1\frac{3}{16}, 1\frac{1}{4}, 1\frac{3}{8}, 1\frac{1}{2}\}$, $c_9 = 100$, $\nu_j|_{\mathcal{U}_9} = \{0, 1, 2, 3, 4, 5\}$.

of accumulated granted points. With a fixed value of $u_8^{\max}$, it is intuitive to conclude from Fig. 5.4 that when incentive is reduced (via the suppression of expected points) the intervention's outcome quality can be impacted.

5.4.2 Multi-phase intervention.

Contrary to a single-phase intervention design, a gradual increase of intervention intensity/dosages over time may be desired in many application settings. In the physical activity behavioral problem, Fig. 5.5 illustrates a multi-step intervention simulation that gradually drives an individual from a *sedentary* baseline state to an *active* state. In this case, the HMPC controller is reconfigured through an online adjustment of $u_8^{\min}$ and $u_8^{\max}$ at predefined stages or instances so that all phase requirements are met.

5.5 Chapter Summary

In this chapter, a more elaborate discussion on the semiphysical identification of a lower-complexity SCT model (OCSE model) was provided, followed by an outline summarizing some of the challenges and limitations arising in the context of *Just Walk*. Next, a participant-validated, semiphysical SCT model identified using techniques from the previous chapters was featured. More specifically, a time series plot and step responses of the OCSE model, highlighting some of the most important individual (i.e., participant) behavioral characteristics, were presented.

Further, the receding horizon strategy used in the the design of closed-loop intervention simulations was introduced; an outline for the main HMPC formulation followed. Application-specific design considerations relevant to the physical activity behavioral problem were provided and analytically expressed using the MLD framework. HMPC-governed closed-loop simulations of the identified SCT-OCSE model

were evaluated with simultaneous setpoint tracking and disturbance rejection; a simple online controller reconfiguration approach was proposed to allow for both single- and multi-phase intervention designs.

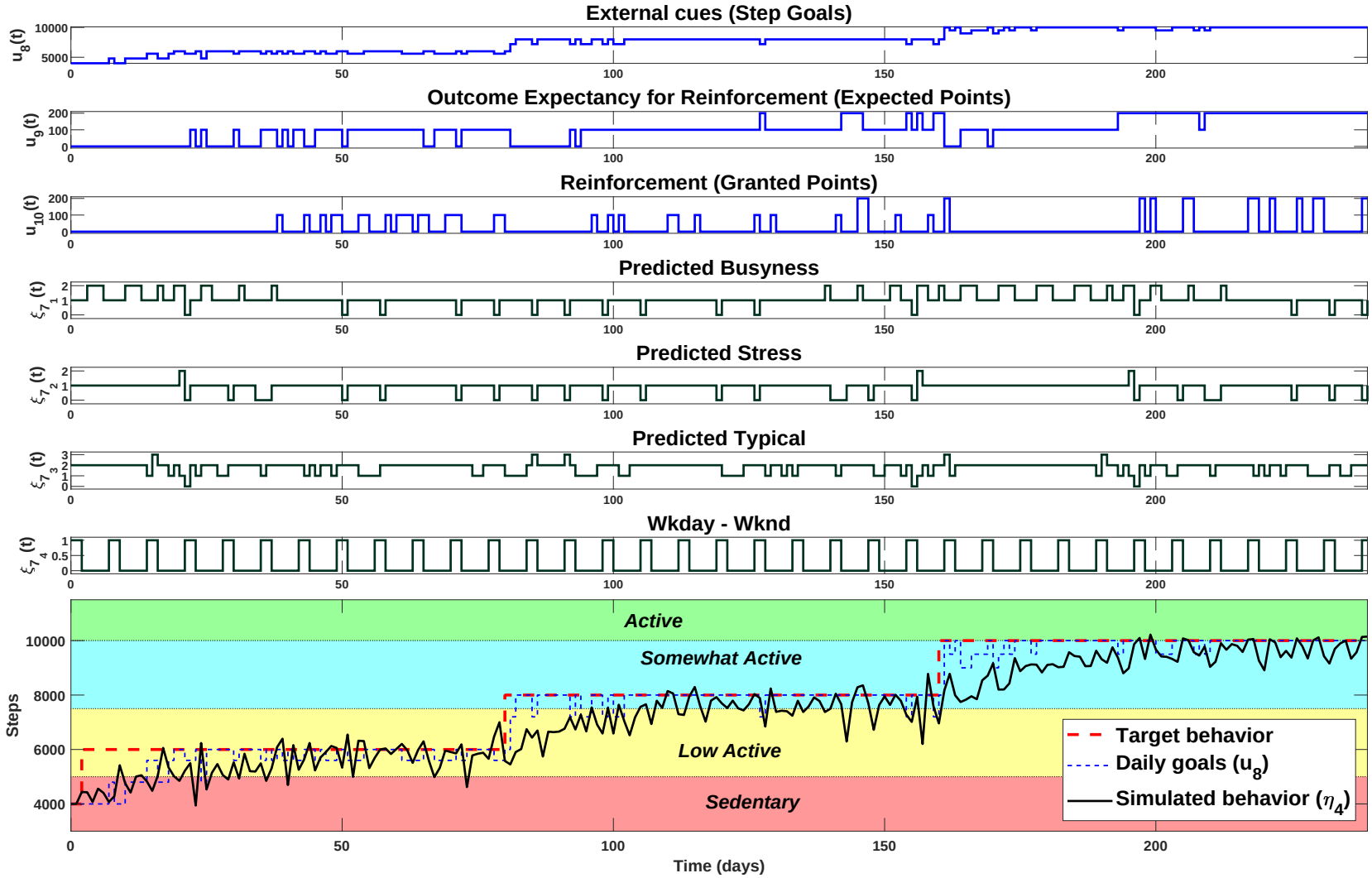

Figure 5.5: HMPC Closed-loop Performance Simulations (Multi-phase Intervention) of a Participant-validated OCSE Model. HMPC Settings Indicated in the Caption of Figure 5.4 Apply with the Following Exceptions: Only the Response Of $Q_u = \text{diag}\{0, 0.1, 0\}$ is simulated; $u_8^{\max} = 6000\ \forall t \in [0, 80]$; $u_8^{\max} = 8000\ \forall t \in [81, 160]$; $u_8^{\max} = 10000\ \forall t \in [161, 240]$.

Chapter 6

SUMMARY, CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Dissertation Summary

This dissertation has continued to demonstrate the viability of using system identification and control systems engineering frameworks in the design of optimized, perpetually adaptive behavioral health interventions. In particular, the work of this dissertation featured the use of real-life, single-subject experimental data in the estimation and validation of both behavioral and energy balance models. While applicable in other domains, the scope of presented work has remained in the domain of designing behavioral health interventions that help prevent or treat behavior-driven and "intergenerational" obesity, conceptually tracing back to intrauterine growth. More specifically, this work was thematically split into two main parts: First, an intergenerational approach included the utilization of the *Healthy Mom Zone* (HMZ) study (providing longitudinal experimental data) in the development, estimation, and validation of a dynamical systems model for regulating infant birth weight was presented in Chapter 2. The *second* part (Chapters 3-5) followed from calls in recent literature for the estimation and validation of dynamic, control-oriented Social Cognitive Theory (SCT) models using longitudinal data from experiments such as *Just Walk* to facilitate the promotion of physical activity among sedentary populations. Of course, the journey of fulfilling these goals has illuminated a number of technical challenges that highlighted and approached in multiple parts of this contribution.

The rest of this chapter provides an executive summary of this dissertation, as well as some conclusions and potential future directions in this research domain.

6.1.1 Intrauterine Fetal Growth Model

Using first principles modeling, particularly the laws of thermodynamics, a first-order, parameter-varying differential equation was developed and estimated in Chapter 2. Following assumptions outlined in Section 2.2.1 and fundamental concepts such as *conservation of energy* and *entropy*, this quasi-LPV model delivers an energy balance weight growth profile as a dynamic function of daily energy intake and expenditure in the forms of maternal dietary intake and physical activity, respectively. A *positivity constraint* was presented to establish the continuous fetal growth *in utero*, providing not only a model validation criterion, but also a potential diagnostic simulation tool for the early detection and prevention of small-for-gestational Age (SGA) and large-for-gestational age (LGA) growth rates. Drawing from existing literature, the presented model includes a modified, intuitive algebraic placental volume equation that can easily be estimated from experimental data guided by prior knowledge.

As opposed to classical cross-sectional studies, the featured HMZ study in this work (Chapter 2) provided a unique opportunity for estimating individual fetoplacental models using longitudinal, single-subject experimental data. Measures and estimates during the second and third trimesters (and birth weight) included daily maternal dietary intake and physical activity, estimated fetal weight and placental volume (from ultrasound measures), fetal body composition (i.e., % body fat), daily glycemic impact of food, and other measures outlined in Symons Downs *et al.*, 2018. The estimated models were validated using a number of validation arguments, including the goodness of fit, contrasting against knowledge from existing literature, and the developed positivity constraint validation criterion. The developed model has been published and cited in recent works (Baller *et al.*, 2019).

6.1.2 System Identification of Just Walk: Social Cognitive Theory Models

As outlined in the introductory chapter, the design of optimized behavioral health interventions using the control system engineering framework typically requires the development of a plant ('human') model; see Figure 1.2. Intuitively, it is clear that a more complete human model must incorporate an *energy balance* component that account for physiological outcomes (e.g., weight gain/loss), as well as a *behavioral* component that enable the prediction (and correction) of the ever-changing psychological states (e.g., self-efficacy). Similar to HMZ, the *Just Walk* pilot study also presented a first-of-a-kind experimental design methodology that was founded on a strong theoretical basis (i.e., Social Cognitive Theory) while simultaneously guided by system identification principles. The *Just Walk* study utilized in this work has also generated longitudinal individual datasets that include input-output measures of behavior and environmental context signals over the course of approximately 14 weeks, posing a unique opportunity for this contribution to estimate and validate pragmatic, control-oriented SCT models. Further, some of the challenges emerging from this pilot *Just Walk* experiment has inspired the development of a new system identification framework for initializing state-of-the-art grey-box solvers in difficult problems.

Using previously published *Just Walk* input-output data, the modeling effort was first segmented into two main consecutive parts: black-box model identification, and grey-box model estimation. With an ultimate goal of estimating and validating control-oriented SCT models, a natural first step is to identify good black-box models using participant data. Thus, individual participant datasets were utilized in establishing input-output causality, as well as in detecting potential input-input correlations (or co-linearity). With an individually identified (i.e., personalized) input-

output dataset for each participant, Auto-Regressive with eXogenous inputs (ARX) structure was considered. The choice of this structure originates from its attractive theoretical properties as further explained in Chapter 3. Next, an exhaustive search approach for the optimal ARX model was the carried out across all possible configurations that include different model orders, input combinations, and estimation/validation data segmentation. A simple penalty weight approach was proposed to underscore models with most favorable statistical properties.

Further, gleaned insights from black-box identification efforts were used to pave the way to semiphysical identification of *Just Walk*. First, a pragmatic, further simplified SCT structure stemming from the semiphysical model in Martín *et al.*, 2016a was introduced and featured as an Operant Conditioning–Self-Efficacy (OCSE) model in Chapter 4. With the absence of prior knowledge of parameter values, combined with the unidentifiable OCSE structure in the context of *Just Walk*, a new method for judicious grey-box solver initialization was proposed as part of this contribution. This Spectral Decomposition (SD) formulation relies on identifying fully-parametrized, physically realizable subspace (black-box) models. Theoretical results include the proposal of a sufficient condition for the existence of a given structure under the proposed formulation; a discussion highlighting sufficient conditions for uniqueness and identifiability followed.

Eigenvalue-constrained subspace models were established and used for the estimation of semiphysical (grey-box) OCSE-SCT models. The estimated OCSE models were validated using the NRMSE goodness of fit criterion over cross-validation data segments, in addition to observations drawn from the classic Bland-Altman (agreement) and residual analysis (correlation) plots. Results from model validation clearly marked the statistical improvements (i.e., bias and variance) as structural, SCT-driven insights are incorporated into the initial crude black-box model.

6.1.3 HMPC Loop Evaluation Using Participant-Validated Models

Following the delivery of participant-validated SCT models, it was possible to accomplish the central goal of this dissertation, which is to demonstrate (using data-driven SCT models) the viability of the dynamical systems and control approach for designing optimized and perpetually adaptive behavioral health interventions. In keeping with the trend of recent works in this domain, Chapter 5 adopted the Hybrid Model Predictive Control (HMPC) framework for adaptive intervention design. A brief overview of the Bemporad HMPC formulation followed by a review of application-specific controller configuration requirements were presented, Finally, a real-life participant-validated model was used to produce and evaluate HMPC-governed closed-loop simulations of single-phase and multi-phase intervention designs. Closed-loop performances, including simultaneous setpoint tracking and (un)measured disturbance rejection, were evaluated under different HMPC tuning settings.

6.2 Conclusions

A number of main conclusions, including conceptual, experimental, and modeling conclusions are outlined below for the interested future researcher. Furthermore, in a following subsection, a brief address to the concerned behavioral science and medicine societies is provided.

- In lieu of using hypothetical models and simulations, the work of this dissertation relied on experimental data drawn from real-life human participants and well-established system identification approaches to demonstrate the efficacy and amenability of the dynamical systems framework in importing and capturing some of the key concepts of the most popular behavioral theories in Psychology.

- Even at the very basic level of correlation analysis and black-box modeling, results from the *Just Walk* did in fact underscore the *idiosyncratic* nature of human behavior, and that a conceivably effective behavioral health intervention must consider the individual environmental context as well as *adapt* to the changing human needs. For example, using the ARX estimator, it was demonstrated that some individuals may be goal-driven, while a "busy" day may predict higher or lower levels of physical activity.

- In a collaboration with the authors of dos Santos *et al.*, 2018, a clear limitation from the pilot *Just Walk* experiment was shown to be that low-frequency characteristics (i.e., steady-state information) were not effectively captured due to insufficient low-frequency excitation, which in turn would require longer and thus more expensive studies of the same sample size. Step and frequency responses of multiple identified MoliZoft models in dos Santos *et al.*, 2018 were plotted to demonstrate this fact; agreement was only found in high frequency (transient responses), with little to no agreement in low frequency (steady-state) responses. While this may be acceptable (to an extent) for a closed-loop intervention design operating on a daily scale, it may be unacceptable for drawing conclusions or issuing predictions at steady-state, which has a value of its own.

  In retrospect, the designer of a *Just Walk*-like experiment may consider the incorporation of Pseudo Random Binary Sequence (PRBS) excitations (with longer pulses) in combination with the used multisine inputs used in Equation 3.1, resulting in a longer experiment at the expense of a lower sample size, or even resorting to a close-loop identification design that can typically lend more tolerance for longer studies. Moreover, longer experiments may be needed

to incorporate more variability in environmental context signals, some of which only produce sufficient variability over longer periods of time. The designer may also want to consider the feasibility of adapting a higher resolution (i.e., a finer time scale) for measuring inputs (intervention dosages), disturbances (environmental context), and output (behavior); it has been analytically shown in Chapter 4 that lower sampling time will naturally result into more accurate estimates of SCT models.

- The process of estimating and validating semiphysical models of Social Cognitive Theory in the context of *Just Walk* has revealed the importance of measuring all possible states, whether directly or via a strong proxy. Unfortunately, in the case of validating the OCSE model of Social Cognitive Theory in the context of Just Walk, daily measurements of *self-efficacy* and *behavioral outcomes* were absent. This has resulted in the structure being unidentifiable, which invited a host of limitations both in model estimation and validation as outlined in Chapters 4 and 5. The interested future researcher may want to consider as many model states as possible to reduce the dimensionality of the estimation problem, or be able to glean more *a priori* knowledge about pivotal model parameter values (as illustrated in Section 4.4.2).

- The HMPC framework remains to provide the most suitable strategy for closed-loop design of optimized, adaptive behavioral health interventions for promoting healthy levels of physical activity. Equipped with the ability to issue optimal hybrid decisions, HMPC control can be configured to incorporate logical, physical, environmental, and financial constraints that define the character of the needed intervention.

6.2.1 Remarks for the Behavioral Science and Medicine Communities

As mentioned in previous chapters, this dissertation work has been motivated by prior efforts in this application domain. The full-blown SCT model was proposed in Martín *et al.*, 2014 noting that "testing these models with actual data is critical" and "[t]he model *structure* must be more deeply validated, via data that [come] from experiments". In their final remarks, Riley *et al.*, 2015 noted that "[o]nly by testing various components of the model with actual data will we be able to determine if this complexity is necessary or if the model can be further simplified and streamlined", concluding that rigorous computational approaches are "needed for health behavior theory testing and intervention development."

The work of Chapters 3 and 4 attempted to explore the quoted remarks in the context of the pilot *Just Walk* study that was designed using a number of principles from system identification. In Chapter 3, some of the most important conclusions drawn from exploring simple, black-box ARX models on "actual data" (i.e., *Just Walk*) pointed that the SCT model can potentially be reduced and further simplified. For example, in addition to reduction by established idiosyncrasy (i.e., personalization of inputs/predictors; Figure 3.4), Table 3.1 and similar efforts with the MoliZoft modeling methods in dos Santos *et al.*, 2018 all point out to the potential viability of a reduction to a third-order SCT model. Based on these insights, the OCSE model "*structure*" in Figure 4.2 was proposed and studied in Chapters 4 and 5.

Now that a further simplified, third-order OCSE model of Social Cognitive Theory is established, it is only natural to estimate and validate such a model in the context of *Just Walk* experimental data. However, there are two types of challenges associated with that: *First*, as discussed in Chapter 4, the prediction-error method (PEM) typically used for estimating grey-box models such as OCSE from input-output data

(used in Martín *et al.*, 2014) is nonconvex and thus principally depends on good initial guesses for parameter values for its success in reaching a global optimum solution or at least a "good enough" local one (see first column "PEM (Poor Initialization)" in Figure 4.3). *Second*, it was shown in Chapter 4 that in the absence of reliable and informative measurements for both *self-efficacy* and *behavioral outcomes and* prior knowledge concerning model parameter values (at least the 'pivotal' ones discussed in Section 4.4.2), the OCSE model becomes fundamentally unidentifiable and cannot be formally validated (or invalidated) using input and *behavior* data. This was also the case with the proposed SCT model in Martín *et al.*, 2014 in the context of the used MILES data. Nonetheless, the effort in Chapter 4 was concerned with proposing this reduced OCSE structure with preliminary model validation using data in Appendix A.

In future *Just Walk*-like experiments, once actual informative and reliable data are available for *all* model states such as *behavior*, *self-efficacy*, and *behavioral outcomes* for the OCSE model in Figure 4.2 (also the *Cue to Action* and *Outcome Expectancy* states for the model in Figure 1.1), the SCT model in question becomes identifiable from data. The contribution in Chapter 4 distilled in Equation (4.15) delivered a convexification of the grey-box estimation problem (see Example 4.3.1), providing a valuable formal model validation (or invalidation) and "theory testing" tool that is 'immune' to potential "local minima" challenges provided that both assumptions in Theorem 4.2.1 are held. When and if a formal OCSE model validation is established in future *Just Walk*-like studies (with *all* model states reliably measured), strategies informed by the various constructs (i.e., *self-efficacy*, etc.) can be developed and utilized in an HMPC-based closed-loop intervention design scheme.

6.3 Potential Future Directions

As the scope of work in this dissertation explored a library of dynamical systems models useful for designing optimized and adaptive behavioral health interventions, an initial effort constrained by experimental limitations; the following are a few interesting directions of potential future work relevant to this research.

6.3.1 Further developments in the fetal growth model

In the work of Chapter 2, we postulated an empirical, 'modified' logistic function characterizing the phases of placental volume growth from conception through birth. Unfortunately, this function (Equation (2.29)) does not incorporate nor produce any further physiological insights other than the general growth profile shaped by the known cell multiplication and spacial constraints. In future developments, it may prove to be useful having a model that characterizes all significant influences on placental growth, other than just *time*. This model can stem either from first principles modeling or by coupling known physiological factors with experimental results. Further, following observations from Thomas *et al.*, 2008, the impact of maternal physical activity on fetal growth was assumed to be mediated by the placental volume, i.e.,

$$\gamma(t) = \alpha PA(t) + \beta \tag{2.22}$$

This needs to be further established using future studies and/or new analyses of potentially existing data; see Baller *et al.*, 2019 for a more elaborate reference of existing studies.

Finally, while it might have been clear to justify the assumed *positivity* (or non-negativity) of fetal growth rate throughout gestation, it is however, not clear in a mathematical sense how to capture events such as starvation or malnutrition. The potential for developing a more complete model in that respect may lie in revealing

the functional form of the $\alpha_W(t)$ in

$$\hat{I}_f(t) = \gamma(t)\left[g(t)m(t) + \alpha_W(t)W_m(t)\right]P(t) \tag{2.14}$$

6.3.2 LPV Modeling of Social Cognitive Theory

While identification results from earlier chapters show real promise for the potency of Linear Time-Invariant (LTI) dynamical systems in predicting and explaining human behavior, it is not only conceivable but in fact known that underlying dynamics driving behavior change can indeed be time-varying and nonlinear (Hayes *et al.*, 2007; Korinek *et al.*, 2018). Given the well-understood theory and properties of linear systems in both identification and control, one may be interested in exploring nonlinearities and time-varying characteristics driving human behavior-change using the LTI framework, which is on offer by LPV identification. Similar to the provided reasoning that resulted in choice of ARX modeling in Chapter 3, and following the general advice from Ljung, 1999 on "*try simple things first!*", an LPV-ARX model structure is thus the recommended choice of initial structure for identification, followed by introducing and exploring practical LPV extensions to the semiphysical OCSE-SCT model. The attractive properties of the LPV-ARX model structure such as *consistency* (Cox, 2018, Theorem 6.1) and *convexity* (existence of a unique analytical solution) continue to extend from LTI-ARX.

6.3.3 Application of HMPC Design in 'Real-life' Intervention Settings

Despite that all models produced in this contribution represent real-life participants from *Just Walk*, one is indeed welcome to question the efficacy of the estimated behavioral models or proposed intervention designs over an extended periods of time not only in the scale of months, but several years. Guo has written a similar section in a recent dissertation relevant to this application domain, stating that "it would

be useful to examine how models could be obtained for the intervention in practice," and that, similar to the case with this dissertation, "HMPC-based control was not performed online." As such, in agreement with Guo, 2018, an adaptive, closed-loop identification strategy is proposed to approach the 'longevity' argument. One may consider such a design in future *Just Walk*-like experiments by starting with an initial model estimated from a short baseline, open-loop dataset or from an "averaged" participant response data from prior analyses as in Chapters 3 and 4.

6.3.4 Toward the 'Convexification' of Semiphysical Identification & Extensions to the SD Method

The proposed Spectral Decomposition (SD) identification algorithm in this work benefited two main assumptions outlined in Section 4.2. Namely, it was assumed that the initially obtained black-box model possesses a number of *distinct* eigenvalues (i.e., poles) that is equal to the order of the *minimal* state-space representation of that system. The second assumption is that input excitations in the data-generating experiment follow a zero-order hold intersampling behavior. A powerful extension may consider to deliver further derivations and/or explicit conditions that enable the elimination of one (or both) of these assumptions.

Additionally, the SD algorithm introduced a loss function that is nonlinear and is thus nonconvex with respect to the similarity transformation T_f. In order to become amenable to the larger family of optimization methods and solvers, one may be interested in deriving exact or at least good gradient and Hessian approximations for solving (4.22).

Further, on a theoretical level, the delivered SD formulation has only considered *linear* grey-box structures. It is noted that one may also consider the extension of delivered theoretical results in Chapter 4 for *quadratic* structures which may be

needed for further model specification. A more general quadratic structure quintuple (P, d, Q_i, q_i, r_i) can result into the following quadratically constrained quadratic program (QCQP)

$$\begin{aligned} \min_{\theta} \quad & \frac{1}{2}\theta^T H\theta + f^T\theta \\ \text{s.t.} \quad & \frac{1}{2}\theta^T Q_i\theta + q_i^T\theta + r_i \leq 0 \quad \forall i = 1, \ldots, l \\ & P\theta = d \end{aligned} \tag{6.1}$$

with variables defined in (4.11) and Q_i, q_i and r_i enforcing the quadratic structural constraints; l is the number of established quintuplets. Note the amenability of this formulation to the incorporation of integer and/or binary constraints. Finally, it is known from Chapter 4 that for an estimation problem with H and Q_i being positive semidefinite (i.e., $H \succeq 0$ and $Q_i \succeq 0$ $\forall i = 1, \ldots, l$), the resulting QCQP in (6.1) remains convex.

Adams, M. A., J. F. Sallis, G. J. Norman, M. F. Hovell, E. B. Hekler and E. Perata, "An adaptive physical activity intervention for overweight adults: A randomized controlled trial", PloS one **8**, 12, e82901 (2013).

Ajzen, I., "The theory of planned behavior", Organizational behavior and human decision processes **50**, 2, 179–211 (1991).

Arleo, E. K., R. N. Troiano, R. da Silva, D. Greenbaum and H. J. Kliman, "Utilizing two-dimensional ultrasound to develop normative curves for estimated placental volume", American Journal of Perinatology **31**, 08, 683–688 (2014).

Azpurua, H., E. F. Funai, L. M. Coraluzzi, L. F. Doherty, I. E. Sasson, M. Kliman and H. J. Kliman, "Determination of placental weight using two-dimensional sonography and volumetric mathematic modeling", American Journal of Perinatology **27**, 02, 151–155 (2010).

Baller, D., D. M. Thomas, K. Cummiskey, C. Bredlau, N. Schwartz, K. Orzechowski, R. C. Miller, A. Odibo, R. Shah and C. M. Salafia, "Gestational growth trajectories derived from a dynamic fetal–placental scaling law", Journal of the Royal Society Interface **16**, 159, 20190417 (2019).

Bandura, A., *Social Foundations of Thought and Action: A Social Cognitive Theory* (Prentice-Hall series in social learning theory, 1986).

Barker, D., *Nutrition in the Womb: How Better Nutrition During Development Will Prevent Heart Disease, Diabetes and Stroke : an Account of the Developmental Origins of Health and Disease, and a Call for Action* (David Barker, 2008).

Bauer, U. E., P. A. Briss, R. A. Goodman and B. A. Bowman, "Prevention of chronic disease in the 21st century: elimination of the leading preventable causes of premature death and disability in the USA", The Lancet **384**, 9937, 45 – 52 (2014).

Bellman, R. and K. J. Åström, "On structural identifiability", Mathematical biosciences **7**, 3-4, 329–339 (1970).

Bemporad, A., "Hybrid Toolbox - User's Guide", `http://cse.lab.imtlucca.it/~bemporad/hybrid/toolbox` (2004).

Bemporad, A. and M. Morari, "Control of systems integrating logic, dynamics, and constraints", Automatica **35**, 3, 407–427 (1999).

Bernstein, I. M. and P. M. Catalano, "Ultrasonographic estimation of fetal body composition for children of diabetic mothers", Investigative Radiology **26**, 8, 722–726 (1991).

Boyd, S. and L. Vandenberghe, *Convex optimization* (Cambridge university press, 2004).

Catalano, P. and H. Ehrenberg, "Review article: The short-and long-term implications of maternal obesity on the mother and her offspring", BJOG: An International Journal of Obstetrics & Gynaecology **113**, 10, 1126–1133 (2006).

Cawley, J. and C. Meyerhoefer, "The medical care costs of obesity: an instrumental variables approach", Journal of health economics **31**, 1, 219–230 (2012).

Çengel, Y. A. and M. A. Boles, *Thermodynamics: An Engineering Approach (Mcgraw-Hill Series in Mechanical Engineering)* (McGraw-Hill Science/Engineering/Math, 2005).

Cetin, I., G. Alvino and M. Cardellicchio, "Long chain fatty acids and dietary fats in fetal nutrition", The Journal of Physiology **587**, 14, 3441–3451, URL `http://dx.doi.org/10.1113/jphysiol.2009.173062` (2009).

Chandler-Laney, P. C. and N. C. Bush, "Maternal obesity, metabolic health, and prenatal programming of offspring obesity", Open Obes J **3**, 42–50 (2011).

Chilali, M. and P. Gahinet, "H_∞ design with pole placement constraints: an lmi approach", IEEE Transactions on automatic control **41**, 3, 358–367 (1996).

Chow, C. C. and K. D. Hall, "The dynamics of human body weight change", PLoS computational biology **4**, 3, e1000045 (2008).

Christiansen, E., L. Garby and T. I. Sørensen, "Quantitative analysis of the energy requirements for development of obesity", Journal of Theoretical Biology **234**, 1, 99 – 106 (2005).

Clapp, J. F., "The effects of maternal exercise on fetal oxygenation and feto-placental growth", European Journal of Obstetrics & Gynecology and Reproductive Biology **110**, S80–S85 (2003).

Cox, P. B., *Towards efficient identification of linear parameter-varying state-space models*, Ph.D. thesis, Eindhoven University of Technology (2018).

Delforge, J., "A sufficient condition for identifiability of a linear system", Mathematical Biosciences **61**, 1, 17–28 (1982).

Demerath, E. W. and D. A. Fields, "Body composition assessment in the infant", American Journal of Human Biology **26**, 3, 291–304 (2014).

Demerath, E. W., W. Johnson, B. A. Davern, C. G. Anderson, J. S. Shenberger, S. Misra and S. E. Ramel, "New body composition reference charts for preterm infants", The American Journal of Clinical Nutrition **105**, 1, 70–77 (2016).

Deshpande, S., N. N. Nandola, D. E. Rivera and J. Younger, "A control engineering approach for designing an optimized treatment plan for fibromyalgia", in "Proceedings of the American Control Conference", pp. 4798 – 4803 (2011).

Deshpande, S., N. N. Nandola, D. E. Rivera and J. W. Younger, "Optimized treatment of fibromyalgia using system identification and hybrid model predictive control", Control Engineering Practice **33**, 1, 161–173 (2014a).

Deshpande, S., D. E. Rivera and J. Younger, "Towards patient-friendly input signal design for optimized pain treatment interventions", Proceedings of the 16th IFAC Symposium on System Identification pp. 1311–1316 (2012).

Deshpande, S., D. E. Rivera, J. W. Younger and N. N. Nandola, "A control systems engineering approach for adaptive behavioral interventions: illustration with a fibromyalgia intervention", Translational behavioral medicine **4**, 3, 275–289 (2014b).

Dong, Y., *A novel control engineering approach to designing and optimizing adaptive sequential behavioral interventions*, Ph.D. thesis, Chemical Engineering, Arizona State University (2014).

Dong, Y., S. Deshpande, D. E. Rivera, D. Symons Downs and J. S. Savage, "Hybrid model predictive control for sequential decision policies in adaptive behavioral interventions", in "Proceedings of the American Control Conference", pp. 4198–4203 (2014).

Dong, Y., D. E. Rivera, D. Symons Downs, J. S. Savage, D. M. Thomas and L. M. Collins, "Hybrid model predictive control for optimizing gestational weight gain behavioral interventions", in "Proceedings of the American Control Conference", pp. 1973 – 1978 (2013).

Dong, Y., D. E. Rivera, D. M. Thomas, J. E. Navarro-Barrientos, D. Symons Downs, J. S. Savage and L. M. Collins, "A dynamical systems model for improving gestational weight gain behavioral interventions", in "Proceedings of the American Control Conference", pp. 4059–4064 (2012).

dos Santos, P. L., M. Freigoun, D. Rivera, E. Hekler, C. Martín, R. Romano, T. Perdicoúlis and J. Ramos, "A molizoft system identification approach of the just walk data", IFAC-PapersOnLine **50**, 1, 12508–12513 (2017).

dos Santos, P. L., M. T. Freigoun, C. A. Martín, D. E. Rivera, E. B. Hekler, R. A. Romano and T. P. Perdicoúlis, "System identification of *Just Walk*: Using matchable-observable linear parametrizations", IEEE Transactions on Control Systems and Technology (2018).

Dudenhausen, J. W., A. Grünebaum and W. Kirschner, "Prepregnancy body weight and gestational weight gain—recommendations and reality in the USA and in Germany", American Journal of Obstetrics & Gynecology **213**, 4, 591–592 (2015).

Ferraro, Z. M., L. Gaudet and K. B. Adamo, "The potential impact of physical activity during pregnancy on maternal and neonatal outcomes", Obstetrical & Gynecological Survey **67**, 2, 99–110 (2012).

Finkelstein, E. A., J. G. Trogdon, J. W. Cohen and W. Dietz, "Annual medical spending attributable to obesity: payer-and service-specific estimates", Health affairs **28**, 5, w822–w831 (2009).

Flegal, K. M., D. F. Williamson, E. R. Pamuk and H. M. Rosenberg, "Estimating deaths attributable to obesity in the united states", American journal of public health **94**, 9, 1486–1489 (2004).

Flood, A., A. F. Subar, S. G. Hull, T. P. Zimmerman, D. J. Jenkins and A. Schatzkin, "Methodology for adding glycemic load values to the National Cancer Institute diet history questionnaire database", Journal of the American Dietetic Association **106**, 3, 393–402 (2006).

Freigoun, M. T., C. A. Martín, A. B. Magann, D. E. Rivera, S. S. Phatak, E. V. Korinek and E. B. Hekler, "System identification of just walk: A behavioral mhealth intervention for promoting physical activity", in "American Control Conference (ACC), 2017", pp. 116–121 (IEEE, 2017).

Freigoun, M. T., D. E. Rivera, P. Guo, E. E. Hohman, A. D. Gernand, D. Symons Downs and J. S. Savage, "A dynamical systems model of intrauterine fetal growth", Mathematical and Computer Modelling of Dynamical Systems **24**, 6, 641–667 (2018).

Freigoun, M. T., K. S. Tsakalis and G. B. Raupp, "A spectral decomposition identification algorithm for structured state-space models: Estimating semiphysical models of social cognitive theory", in "American Control Conference (ACC), 2021", pp. 2830–2835 (IEEE, 2021b).

Freigoun, M. T., K. S. Tsakalis and G. B. Raupp, "On the identification of social cognitive theory models and closed-loop intervention simulations using hybrid model predictive control", (Accepted for publication in Proceedings of the 19$^{\text{th}}$ IFAC Symposium on System Identification, 2021a).

Gallino, V., C. A. Martín, D. Sosa and M. Barcos-Arias, "A control engineering and system identification approach to improve a cocoa biofertilization process", in "2018 Annual American Control Conference (ACC)", pp. 380–385 (2018).

Guillaume, P., J. Schoukens, R. Pintelon and I. Kollar, "Crest-factor minimization using nonlinear Chebyshev approximation methods", IEEE Transactions on Instrumentation and Measurement **40**, 6, 982–989 (1991).

Guo, P., *System Identification, State Estimation, And Control Approaches to Gestational Weight Gain Interventions*, Ph.D. thesis (2018).

Guo, P., D. E. Rivera, D. Symons Downs and J. S. Savage, "Semi-physical identification and state estimation of energy intake for interventions to manage gestational weight gain", in "2016 American Control Conference (ACC)", pp. 1271–1276 (2016).

Guo, P., D. E. Rivera, J. S. Savage and D. S. Downs, "State estimation under correlated partial measurement losses: Implications for weight control interventions", IFAC-PapersOnLine **50**, 1, 13532–13537 (2017).

Guo, P., D. E. Rivera, J. S. Savage, E. E. Hohman, A. M. Pauley, K. S. Leonard and D. S. Downs, "System identification approaches for energy intake estimation: Enhancing interventions for managing gestational weight gain", IEEE Transactions on Control Systems Technology (2018).

Hadlock, F., R. Harrist, R. Carpenter, R. Deter and S. Park, "Sonographic estimation of fetal weight. the value of femur length in addition to head and abdomen measurements.", Radiology **150**, 2, 535–540 (1984).

Hadlock, F. P., R. Harrist, R. S. Sharman, R. L. Deter and S. K. Park, "Estimation of fetal weight with the use of head, body, and femur measurements—a prospective study", American Journal of Obstetrics & Gynecology **151**, 3, 333–337 (1985).

Hadlock, F. P., R. B. Harrist and J. Martinez-Poyer, "In utero analysis of fetal growth: a sonographic weight standard.", Radiology **181**, 1, 129–133 (1991).

Hall, K. D., "Predicting metabolic adaptation, body weight change, and energy intake in humans", American Journal of Physiology-Endocrinology and Metabolism **298**, 3, E449–E466 (2009).

Hall, K. D., "Mechanisms of metabolic fuel selection: modeling human metabolism and body-weight change", IEEE Engineering in Medicine and Biology Magazine **29**, 1, 36–41 (2010).

Hall, K. D., "Estimating human energy intake using mathematical models", The American Journal of Clinical Nutrition **100**, 3, 744–745 (2014).

Hall, K. D. and P. N. Jordan, "Modeling weight-loss maintenance to help prevent body weight regain", The American journal of clinical nutrition **88**, 6, 1495–1503 (2008).

Haskell, W. L., S. N. Blair and J. O. Hill, "Physical activity: Health outcomes and importance for public health policy", Preventive Medicine **49**, 4, 280 – 282 (2009).

Hayes, A. M., J.-P. Laurenceau, G. Feldman, J. L. Strauss and L. Cardaciotto, "Change is not always linear: The study of nonlinear and discontinuous patterns of change in psychotherapy", Clinical Psychology Review **27**, 6, 715–723 (2007).

Hekler, E. B., "Just walk study", `http://justwalkstudy.weebly.com/`, [Online; accessed September-23-2015] (2015).

Hekler, E. B., D. E. Rivera, C. A. Martin, S. S. Phatak, M. T. Freigoun, E. Korinek, P. Klasnja, M. A. Adams and M. P. Buman, "Tutorial for using control systems engineering to optimize adaptive mobile health interventions", Journal of Medical Internet Research **20**, 6, URL `https://www.jmir.org/2018/6/e214` (2018).

Hillier, T. A., K. L. Pedula, M. M. Schmidt, J. A. Mullen, M.-A. Charles and D. J. Pettitt, "Childhood obesity and metabolic imprinting", Diabetes Care **30**, 9, 2287–2292 (2007).

J. Löfberg, "YALMIP : a toolbox for modeling and optimization in MATLAB", Proceedings of the IEEE International Symposium on Computer Aided Control Systems Design pp. 284–289 (2004).

Jinyuan, L., T. Wan, C. Guanqin, L. Yin, F. Changyong *et al.*, "Correlation and agreement: overview and clarification of competing concepts and measures", Shanghai archives of psychiatry **28**, 2, 115 (2016).

Kane, M. B., "Modeling human-in-the-loop behavior and interactions with hvac systems", in "2018 Annual American Control Conference (ACC)", pp. 4628–4633 (IEEE, 2018).

Kennaugh, J. M. and W. W. Hay Jr, "Nutrition of the fetus and newborn", Western Journal of Medicine **147**, 4, 435 (1987).

Keys, A., J. Brožek, A. Henschel, O. Mickelsen and H. L. Taylor, "The biology of human starvation.(2 vols).", (1950).

King, A. C., E. B. Hekler, L. A. Grieco, S. J. Winter, J. L. Sheats, M. P. Buman, B. Banerjee, T. N. Robinson and J. Cirimele, "Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults", PLoS ONE **8**, 4, e62613 (2013).

Klein, R., "Bland-altman and correlation plot", Mathworks File Exchange (2014).

Korinek, E. V., S. S. Phatak, C. A. Martin, M. T. Freigoun, D. E. Rivera, M. A. Adams, P. Klasnja, M. P. Buman and E. B. Hekler, "Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention", Journal of behavioral medicine **41**, 1, 74–86 (2018).

Langhoff-Roos, J., G. Lindmark and M. Gebre-Medhin, "Maternal fat stores and fat accretion during pregnancy in relation to infant birthweight", Br J Obstet Gynaecol **94**, 1170–1177 (1987).

Lindskog, P., *Methods, algorithms and tools for system identification based on prior knowledge*, Ph.D. thesis, Linköping University (1996).

Lindskog, P. and L. Ljung, "Tools for semiphysical modelling", International Journal of Adaptive Control and Signal Processing **9**, 6, 509–523 (1995).

Ljung, L., "From data to model: A guided tour", in "Control, 1994. Control'94. International Conference on", vol. 1, pp. 422–430 (IET, 1994).

Ljung, L., *System Identification: Theory for the User* (Upper Saddle River, NJ: Prentice Hall PTR, 1999), 2nd edn.

Ljung, L., "On convexification of system identification criteria", Automation and Remote Control **80**, 9, 1591–1606 (2019).

Martín, C. A., *A System Identification and Control Engineering Approach for Optimizing mHealth Behavioral Interventions Based on Social Cognitive Theory*, Ph.D. thesis, Electrical Engineering, Arizona State University (2016).

Martín, C. A., D. E. Rivera and E. B. Hekler, "Design of informative identification experiments for behavioral interventions", in "Proceedings of the 17th IFAC Symposium on System Identification", pp. 1325–1330 (2015).

Martín, C. A., D. E. Rivera and E. B. Hekler, "A decision framework for an adaptive behavioral intervention for physical activity using hybrid model predictive control", in "Proceedings of the American Control Conference", pp. 3576–3581 (2016a).

Martín, C. A., D. E. Rivera and E. B. Hekler, "An enhanced identification test monitoring procedure for MIMO systems relying on uncertainty estimates", in "Submitted to the 55th IEEE Conference on Decision and Control", (2016b).

Martín, C. A., D. E. Rivera, W. T. Riley, E. B. Hekler, M. P. Buman, M. A. Adams and A. C. King, "A dynamical systems model of Social Cognitive Theory", in "Proceedings of the American Control Conference", pp. 2407–2412 (2014).

Symons Downs, D., J. S. Savage, D. E. Rivera, J. M. Smyth, B. J. Rolls, E. E. Hohman, K. M. McNitt, A. R. Kunselman, C. Stetter, A. M. Pauley, K. S. Leonard and P. Guo, "Individually tailored, adaptive intervention to manage gestational weight gain: Protocol for a randomized controlled trial in women with overweight and obesity", JMIR Research Protocols **7**, 6, e150, URL `https://doi.org/10.2196/resprot.9220` (2018).

McTiernan, A., "Mechanisms linking physical activity with cancer", Nature Reviews Cancer **8**, 3, 205–211 (2008).

Mercere, G., "Technical committee on system identification and adaptive control [technical activities]", IEEE Control Systems Magazine **37**, 4, 13–16 (2017).

Mercère, G., O. Prot and J. A. Ramos, "Identification of parameterized gray-box state-space systems: From a black-box linear time-invariant representation to a structured one", IEEE Transactions on Automatic Control **59**, 11, 2873–2885 (2014).

Miller, D., R. de Callafon and M. Brenner, "Eigenvalue constraints for realization-based identification", in "AIAA Atmospheric Flight Mechanics Conference", p. 4951 (2012).

Miller, D. N. and R. A. De Callafon, "Subspace identification with eigenvalue constraints", Automatica **49**, 8, 2468–2473 (2013).

Morari, M. and E. Zafiriou, *Robust Process Control* (Prentice-Hall International, 1989).

Mu, J., J. C. Slevin, D. Qu, S. McCormick and S. L. Adamson, "In vivo quantification of embryonic and placental growth during gestation in mice using micro-ultrasound", Reproductive Biology and Endocrinology **6**, 1, 34 (2008).

Nandola, N. N. and D. E. Rivera, "An improved formulation of hybrid model predictive control with application to production-inventory systems", IEEE Transactions on Control Systems Technology **21**, 1, 121–135 (2013).

Navarro-Barrientos, J. E., D. E. Rivera and L. M. Collins, "A dynamical model for describing behavioural interventions for weight loss and body composition change", Mathematical and Computer Modelling of Dynamical Systems **17**, 2, 183–203 (2011).

Nilsen, W., E. Ertin, E. B. Hekler, S. Kumar, I. Lee, R. Mangharam, M. Pavel, J. M. Rehg, W. Riley, D. E. Rivera *et al.*, "Modeling opportunities in mhealth cyber-physical systems", in "Mobile Health", pp. 443–453 (Springer, 2017).

Noblet, J., C. Karege, S. Dubois and J. Van Milgen, "Metabolic utilization of energy and maintenance requirements in growing pigs: effects of sex and genotype.", Journal of Animal Science **77**, 5, 1208–1216 (1999).

O'Connor, D. P., M. T. Mahar, M. S. Laughlin and A. S. Jackson, "The bland-altman method should not be used in regression cross-validation studies", Research Quarterly for Exercise and Sport **82**, 4, 610–616 (2011).

Ogunnaike, B. A. and W. H. Ray, *Process Dynamics, Modeling, and Control* (New York : Oxford University Press, 1994).

Orzechowski, K., D. Thomas, C. J. McNamara and R. C. Miller, "Volumetric assessment of longitudinal placental growth.", Obstetrics & Gynecology **123**, 164S (2014).

Owen, N., A. Bauman and W. Brown, "Too much sitting: a novel and important predictor of chronic disease risk?", British Journal of Sports Medicine **43**, 2, 81–83 (2008).

O'Neill, K. A., M. F. Murphy, K. J. Bunch, S. E. Puumala, S. E. Carozza, E. J. Chow, B. A. Mueller, C. C. McLaughlin, P. Reynolds, T. J. Vincent *et al.*, "Infant birthweight and risk of childhood cancer: international population-based case control studies of 40,000 cases", International Journal of Epidemiology **44**, 1, 153–168 (2015).

Parrilo, P. and L. Ljung, *Initialization of physical parameter estimates* (Linköping University Electronic Press, 2003).

Pauley, A. M., E. Hohman, J. S. Savage, D. E. Rivera, P. Guo, K. S. Leonard and D. Symons Downs, "Gestational weight gain intervention impacts determinants of healthy eating and exercise in overweight/obese pregnant women", Journal of obesity **2018** (2018).

Phatak, S. S., M. T. Freigoun, C. A. Martín, D. E. Rivera, E. V. Korinek, M. A. Adams, M. P. Buman, P. Klasnja and E. B. Hekler, "Modeling individual differences: A case study of the application of system identification for personalizing a physical activity intervention", Journal of biomedical informatics **79**, 82–97 (2018).

Phatak, S. S., E. B. Hekler, D. E. Rivera, C. A. Martín and M. T. Freigoun, "Building a dynamical model to predict "ambitious but doable" daily step goals", in "Annual Meeting of the Society of Behavioral Medicine", (United States, Washington DC, 2016).

Pitkin, R. M., "Nutritional support in obstetrics and gynecology", Clinical Bbstetrics and Gynecology **19**, 3, 489–513 (1976).

Qiao, Y., J. Ma, Y. Wang, W. Li, P. T. Katzmarzyk, J.-P. Chaput, M. Fogelholm, W. D. Johnson, R. Kuriyan, A. Kurpad *et al.*, "Birth weight and childhood obesity: a 12-country study", International Journal of Obesity Supplements **5**, S74–S79 (2015).

Regnault, T. R., S. W. Limesand and W. W. Hay Jr, "Factors influencing fetal growth", NeoReviews **2**, 6, e119–e128 (2001).

Rich-Edwards, J. W., M. J. Stampfer, J. E. Manson, B. Rosner, S. E. Hankinson, G. A. Colditz, C. H. Hennekens and W. C. Willet, "Birth weight and risk of cardiovascular disease in a cohort of women followed up since 1976", BMJ **315**, 7105, 396–400 (1997).

Riley, W. T., C. A. Martin, D. E. Rivera, E. B. Hekler, M. A. Adams, M. P. Buman, M. Pavel and A. C. King, "Development of a dynamic computational model of social cognitive theory", Translational behavioral medicine **6**, 4, 483–495 (2015).

Riley, W. T., D. E. Rivera, A. A. Autienza, W. Nilsen, S. Allison and R. Mermelstein, "Health behavior models in the age of mobile interventions: are our theories up to the task?", Translational Behavioral Medicine: Practice, Policy, Research **1**, 1, 53–71 (2011).

Rivera, D. E., E. B. Hekler, J. S. Savage and D. Symons Downs, "Intensively adaptive interventions using control systems engineering: Two illustrative examples", in "Optimization of Behavioral, Biobehavioral, and Biomedical Interventions: Advanced Topics", edited by L. M. Collins and K. C. Kugler, pp. 121–173 (Springer, 2018).

Rivera, D. E., H. Lee, H. D. Mittelmann and M. W. Braun, "Constrained multisine input signals for plant-friendly identification of chemical process systems", Journal of Process Control **19**, 4, 623–635 (2009).

Rivera, D. E., C. A. Martín, K. P. Timms, S. Deshpande, N. N. Nandola and E. B. Hekler, "Control systems engineering for optimizing behavioral mHealth interventions", in "Mobile Health. Sensors, Analytic Methods, and Applications", edited by J. M. Rehg, S. A. Murphy and S. Kumar, pp. 455 – 493 (Springer International Publishing, 2017).

Rivera, D. E., M. D. Pew and L. M. Collins, "Using engineering control principles to inform the design of adaptive interventions: a conceptual introduction", Drug and Alcohol Dependence **88**, 2, S31–S40 (2007).

Savage, J. S., D. Symons Downs, Y. Dong and D. E. Rivera, "Control systems engineering for optimizing a prenatal weight gain intervention to regulate infant birth weight", American Journal of Public Health **104**, 7, 1247–1254 (2014).

Schwartz, J. and H. Galan, "Ultrasound in assessment of fetal growth disorders: is there a role for subcutaneous measurements?", Ultrasound in Obstetrics & Gynecology **22**, 4, 329–335 (2003).

Schwartz, J. D., W. Wang and D. E. Rivera, "Simulation-based optimization of process control policies for inventory management in supply chains", Automatica **42**, 8, 1311–1320 (2006).

Shiffman, S., A. A. Stone and M. R. Hufford, "Ecological momentary assessment", Clinical Psychology **4**, 1, 1–32 (2008).

Spracklen, C. N., R. B. Wallace, S. Sealy-Jefferson, J. G. Robinson, J. L. Freudenheim, M. F. Wellons, A. F. Saftlas, L. G. Snetselaar, J. E. Manson, L. Hou *et al.*, "Birth weight and subsequent risk of cancer", Cancer Epidemiology **38**, 5, 538–543 (2014).

Thomas, D. M., J. F. Clapp and S. Shernce, "A foetal energy balance equation based on maternal exercise and diet", Journal of The Royal Society Interface **5**, 21, 449–455 (2008).

Thomas, D. M., J. E. Navarro-Barrientos, D. E. Rivera, S. B. Heymsfield, C. Bredlau, L. M. Redman, C. K. Martin, S. A. Lederman, L. M. Collins and N. F. Butte, "Dynamic energy-balance model predicting gestational weight gain", The American Journal of Clinical Nutrition **95**, 1, 115–122 (2012).

Thompson, J., L. Irgens, R. Skjaerven and S. Rasmussen, "Placenta weight percentile curves for singleton deliveries", BJOG: An International Journal of Obstetrics & Gynaecology **114**, 6, 715–720 (2007).

Thune, I., T. Brenn, E. Lund and M. Gaard, "Physical activity and the risk of breast cancer", New England Journal of Medicine **336**, 18, 1269–1275, pMID: 9113929 (1997).

Timms, K. P., C. A. Martín, D. E. Rivera, E. B. Hekler and W. T. Riley, "Leveraging intensive longitudinal data to better understand health behaviors", in "Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society", pp. 6888 – 6891 (2014a).

Timms, K. P., D. E. Rivera, L. M. Collins and M. E. Piper, "Understanding and optimizing smoking cessation interventions", in "Proceedings of the American Control Conference", pp. 1967–1972 (2013).

Timms, K. P., D. E. Rivera, L. M. Collins and M. E. Piper, "Continuous-time system identification of a smoking cessation intervention,", International Journal of Control **87**, 7, 1423–1437 (2014b).

Timms, K. P., D. E. Rivera, L. M. Collins and M. E. Piper, "A dynamical systems approach to understanding self-regulation in smoking cessation behavior change", Nicotine and Tobacco Research, Special Issue on New Methods for Advancing Research on Tobacco Dependence Using Ecological Momentary Assessments **16**, Suppl 2, S159–S168 (2014c).

Timms, K. P., D. E. Rivera, M. E. Piper and L. M. Collins, "A hybrid model predictive control strategy for optimizing a smoking cessation intervention", in "Proceedings of the American Control Conference", pp. 2389 – 2394 (2014d).

Torrisi, F. D. and A. Bemporad, "HYSDEL–a tool for generating computational hybrid models for analysis and synthesis problems", IEEE Transactions on Control Systems Technology **12**, 2, 235–249 (2004).

Trogdon, J., E. A. Finkelstein, T. Hylands, P. Dellea and S. Kamal-Bahl, "Indirect costs of obesity: a review of the current literature", Obesity Reviews **9**, 5, 489–500 (2008).

Troiano, R. P., D. Berrigan, K. W. Dodd, L. C. Mâsse, T. Tilert and M. Mcdowell, "Physical activity in the United States measured by accelerometer", Medicine & Science in Sports & Exercise **40**, 1, 181–188 (2008).

Tudor-Locke, C. and D. R. Bassett, "How many steps/day are enough?", Sports medicine **34**, 1, 1–8 (2004).

Ugray, Z., L. Lasdon, J. Plummer, F. Glover, J. Kelly and R. Martí, "Scatter search and local nlp solvers: A multistart framework for global optimization", INFORMS Journal on Computing **19**, 3, 328–340 (2007).

U.S. Department of Health and Human Services, "2008 Physical Activity Guidelines for Americans", URL `https://health.gov/paguidelines` (2008).

Van Overschee, P. and B. De Moor, *Subspace identification for linear systems: Theory—Implementation—Applications* (Springer Science & Business Media, 2012).

Wallace, J., S. Bhattacharya and G. Horgan, "Gestational age, gender and parity specific centile charts for placental weight for singleton deliveries in Aberdeen, UK", Placenta **34**, 3, 269–274 (2013).

Widdowson, E. M. and C. M. Spray, "Chemical development in utero", Archives of Disease in Childhood **26**, 127, 205 (1951).

Wilmot, E. G., C. L. Edwardson, F. A. Achana, M. J. Davies, T. Gorely, L. J. Gray, K. Khunti, T. Yates and S. J. Biddle, "Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis", (2012).

Yu, C., L. Ljung and M. Verhaegen, "Identification of structured state-space models", Automatica **90**, 54–61 (2018).

Yu, C., L. Ljung, A. Wills and M. Verhaegen, "Constrained subspace method for the identification of structured state-space models", IEEE Transactions on Automatic Control pp. 1–1 (2019).

APPENDIX A

PUBLISHED JUST WALK DATA DIGITIZATION

A.1 Overview

In Chapters 2 and 3, this dissertation had access to the original datasets from the *Healthy Mom Zone* and *Just Walk* studies, respectively. However, for unrestricted publication purposes, the rest of this dissertation strictly resorted to recovering the *Just Walk* data used in the completion of Chapters 4 and 5 from published graphical sources. The purpose of this appendix is to describe the data digitization process in support of results in Chapters 4 and 5 and provide the necessary information and arguments that underline the estimated accuracy and confidence in the obtained final sets.

The primary concern in Chapter 4 was to deliver an identification formulation for estimating structured state-space (grey-box) models such as the Operant Conditioning–Self-Efficacy (OCSE) model introduced in Figure 4.2. In order to preliminarily validate the OCSE model using experimental data of real human participants as well as test the efficacy of the developed spectral decomposition formulation relative to "off-the-shelf" software and methods, input-output data for Participants A and B published in Mercere, 2017 and Freigoun *et al.*, 2017 (respectively) were digitized and recovered with high 'accuracy' (and full confidence in 13 out of 14 total signals; see Table A.7). To promote further inquiry in this area, the final recovered sets from the cited sources are given in Tables A.2-A.6 for the interested researcher.

A.2 Known Facts & Graphical Sets

To support statements of confidence given in Table A.7, the following *Just Walk* information must be highlighted:

1. Recovery of data points from the published *electronic* versions of Mercere, 2017; Freigoun *et al.*, 2017 was carried out using the Adobe Illustrator vector graphics editor, producing descaled data with a precision of 7 decimal points (see Graphical Sets 1-15). Fortunately, both cited sources provided the concerned plots with high accuracy information in the form of `line`, `polyline`, and `polygon` graphical objects that maintain the relative distance between all data points established by the published plot.

   Both `line` and `polyline` graphical objects contain 'unique' descaled data points with precision of 7 decimal places. The one and only case containing a `polygon` object was treated by 'cutting' the `polygon` to create two `polyline` sets (see Graphical Sets 6 and 7). It is known from these objects that the 'true' value must lie within the maximum possible gap between the two `polyline` sets originating from the original `polygon` object.

2. In both signals, *Goals* and *Expected Points*, it is known from Freigoun *et al.*, 2017; Phatak *et al.*, 2018; Korinek *et al.*, 2018 that orthogonal-in-frequency, 16-day repeating cycles were used in the design of these signals (see Table 1 in Korinek *et al.*, 2018 for a real representative sample). Using the known "IF-THEN" rule, the *Granted Points* signal value at day k must be either zero or identical to the *Expected Points* value at day $k-1$ (see Equation 5.3).

3. It is also known from Korinek *et al.*, 2018; Phatak *et al.*, 2018 (which was also visually and numerically verified) that *Expected Points* are issued on a 100-500

point scale, restricted to multiples of 25 only (see Table 1 in Korinek *et al.*, 2018 for a real representative sample).

4. By definition, all other signal values are strictly whole numbers (i.e., 0 or positive integers) on predefined scales in Freigoun *et al.*, 2017; Korinek *et al.*, 2018; Phatak *et al.*, 2018 as follows. *Behavior*: whole number; *Goals*: whole number; *Predicted Busyness*: 1-4; *Predicted Stress*: 1-5; *Predicted Typical*: 1-4; *Weekday - Weekend*: 0 for weekday, 1 for weekend.

5. The total number of available data points in both sources are 88 (for 88 days), starting from a time index of 0. This is verified visually as well as can be inferred from the retrieved Graphical Sets 1-15.

The following Graphical Sets 1-15 include retrieved data points to arbitrary scales from Mercere, 2017; Freigoun *et al.*, 2017 using a high-precision software.

Graphical Set 1: Participant A (Goals)


```
<line id="XMLID_7_" class="st0" x1="3.7169952" y1="10.2731934" x2="0" y2="10.2731934"/>
<line id="XMLID_5_" class="st0" x1="3.7169952" y1="2.6628113" x2="0" y2="2.6628113"/>
<polyline id="XMLID_2_" class="st0" points="318.2929688,2.5610352 318.2929688,16.6767578 314.6357422,16.6767578
      314.6357422,10.9780273 310.9819336,10.9780273 310.9819336,0.3500977 307.3261719,0.3500977
      307.3261719,2.4350586 303.6699219,2.4350586 303.6699219,0.8598633 300.0131836,0.8598633
      300.0131836,14.7709961 296.3579102,14.7709961 296.3579102,13.34375 292.7041016,13.34375
      292.7041016,2.8608398 289.0478516,2.8608398 289.0478516,16.59375 285.3920898,16.59375
      285.3920898,11.0761719 281.7358398,11.0761719 281.7358398,0.25 278.0810547,0.25 278.0810547,2.5180664
      274.4248047,2.5180664 274.4248047,0.7910156 270.7680664,0.7910156 270.7680664,14.7900391
      267.1118164,14.7900391 267.1118164,13.3588867 263.4570313,13.3588867 263.4570313,2.8110352
      259.8017578,2.8110352 259.8017578,16.6767578 256.1469727,16.6767578 256.1469727,10.9780273
      252.4912109,10.9780273 252.4912109,0.3500977 248.8349609,0.3500977 248.8349609,2.4350586
      245.1791992,2.4350586 245.1791992,0.8598633 241.5239258,0.8598633 241.5239258,14.7709961
      237.8691406,14.7709961 237.8691406,13.34375 234.2138672,13.34375 234.2138672,2.8608398
      230.5571289,2.8608398 230.5571289,16.59375 226.9018555,16.59375 226.9018555,11.0761719
      223.2470703,11.0761719 223.2470703,0.25 219.5917969,0.25 219.5917969,2.5180664 215.934082,2.5180664
      215.934082,0.7910156 212.277832,0.7910156 212.277832,14.7900391 208.6240234,14.7900391
      208.6240234,13.3588867 204.9677734,13.3588867 204.9677734,2.8110352 201.3110352,2.8110352
      201.3110352,16.6767578 197.6547852,16.6767578 197.6547852,10.9780273 194,10.9780273 194,0.3500977
      190.34375,0.3500977 190.34375,2.4350586 186.6879883,2.4350586 186.6879883,0.8598633 183.0341797,0.8598633
      183.0341797,14.7709961 179.3779297,14.7709961 179.3779297,13.34375 175.7211914,13.34375
      175.7211914,2.8608398 172.065918,2.8608398 172.065918,16.59375 168.4101563,16.59375 168.4101563,11.0761719
      164.7539063,11.0761719 164.7539063,0.25 161.0981445,0.25 161.0981445,2.5180664 157.4418945,2.5180664
      157.4418945,0.7910156 153.7890625,0.7910156 153.7890625,14.7900391 150.1318359,14.7900391
      150.1318359,13.3588867 146.4760742,13.3588867 146.4760742,2.8110352 142.8208008,2.8110352
      142.8208008,16.6767578 139.1640625,16.6767578 139.1640625,10.9780273 135.5087891,10.9780273
      135.5087891,0.3500977 131.8540039,0.3500977 131.8540039,2.4350586 128.1992188,2.4350586
      128.1992188,0.8598633 124.5429688,0.8598633 124.5429688,14.7709961 120.8857422,14.7709961
      120.8857422,13.34375 117.230957,13.34375 117.230957,2.8608398 113.5761719,2.8608398 113.5761719,16.59375
      109.9199219,16.59375 109.9199219,11.0761719 106.2651367,11.0761719 106.2651367,0.25 102.6079102,0.25
      102.6079102,2.5180664 98.953125,2.5180664 98.953125,0.7910156 95.296875,0.7910156 95.296875,14.7900391
      91.6401367,14.7900391 91.6401367,13.3588867 87.9848633,13.3588867 87.9848633,2.8110352 84.3310547,2.8110352
       84.3310547,16.6767578 80.6757813,16.6767578 80.6757813,10.9780273 77.019043,10.9780273 77.019043,0.3500977
       73.362793,0.3500977 73.362793,2.4350586 69.7080078,2.4350586 69.7080078,0.8598633 66.0507813,0.8598633
      66.0507813,14.7709961 62.3959961,14.7709961 62.3959961,13.34375 58.7412109,13.34375 58.7412109,2.8608398
      55.0849609,2.8608398 55.0849609,16.59375 51.4291992,16.59375 51.4291992,11.0761719 47.7739258,11.0761719
      47.7739258,0.25 44.1181641,0.25 44.1181641,2.5180664 40.4628906,2.5180664 40.4628906,0.7910156
      36.8071289,0.7910156 36.8071289,14.7900391 33.1518555,14.7900391 33.1518555,13.3588867
      29.4960938,13.3588867 29.4960938,2.8110352 25.8398438,2.8110352 25.8398438,16.6767578 22.1850586,16.6767578
       22.1850586,10.9780273 18.5288086,10.9780273 18.5288086,0.3500977 14.8720703,0.3500977 14.8720703,2.4350586
       11.2158203,2.4350586 11.2158203,0.8598633 7.5610352,0.8598633 7.5610352,14.7709961 3.9057617,14.7709961
      3.9057617,13.34375 0,13.34375"/>
```

Graphical Set 2: Participant A (Expected Points)

```
<line id="XMLID_3_" class="st0" x1="3.7169952" y1="9.677887" x2="0" y2="9.677887"/>
<line id="XMLID_2_" class="st0" x1="3.7169952" y1="3.3927002" x2="0" y2="3.3927002"/>
<polyline id="XMLID_4_" class="st0" points="318.2929688,2.8569336 318.2929688,1.0351563 314.637207,1.0351563
      314.637207,1.8208008 310.9819336,1.8208008 310.9819336,11.25 307.3261719,11.25 307.3261719,9.6782227
      303.6699219,9.6782227 303.6699219,1.8208008 300.0131836,1.8208008 300.0131836,6.5361328
      296.3579102,6.5361328 296.3579102,12.8208008 292.7041016,12.8208008 292.7041016,10.4628906
      289.0478516,10.4628906 289.0478516,12.0341797 285.3920898,12.0341797 285.3920898,11.25 281.7358398,11.25
      281.7358398,1.8208008 278.0810547,1.8208008 278.0810547,3.3930664 274.4248047,3.3930664 274.4248047,11.25
      270.7680664,11.25 270.7680664,6.5361328 267.1118164,6.5361328 267.1118164,0.25 263.4570313,0.25
      263.4570313,2.6069336 259.8007813,2.6069336 259.8017578,1.0351563 256.1469727,1.0351563
      256.1469727,1.8208008 252.4912109,1.8208008 252.4912109,11.25 248.8349609,11.25 248.8349609,9.6782227
      245.1791992,9.6782227 245.1791992,1.8208008 241.5239258,1.8208008 241.5239258,6.5361328
      237.8691406,6.5361328 237.8691406,12.8208008 234.2138672,12.8208008 234.2138672,10.4628906
      230.5571289,10.4628906 230.5571289,12.0341797 226.9018555,12.0341797 226.9018555,11.25 223.2470703,11.25
      223.2470703,1.8208008 219.5917969,1.8208008 219.5917969,3.3930664 215.934082,3.3930664 215.934082,11.25
      212.277832,11.25 212.277832,6.5361328 208.6240234,6.5361328 208.6240234,0.25 204.9677734,0.25
      204.9677734,2.6069336 201.3110352,2.6069336 201.3110352,1.0351563 197.6547852,1.0351563
      197.6547852,1.8208008 194,1.8208008 194,11.25 190.34375,11.25 190.34375,9.6782227 186.6879883,9.6782227
      186.6879883,1.8208008 183.0341797,1.8208008 183.0341797,6.5361328 179.3779297,6.5361328
      179.3779297,12.8208008 175.7211914,12.8208008 175.7211914,10.4628906 172.065918,10.4628906
      172.065918,12.0341797 168.4101563,12.0341797 168.4101563,11.25 164.7539063,11.25 164.7539063,1.8208008
      161.0981445,1.8208008 161.0981445,3.3930664 157.4418945,3.3930664 157.4418945,11.25 153.7890625,11.25
      153.7890625,6.5361328 150.1318359,6.5361328 150.1318359,0.25 146.4760742,0.25 146.4760742,2.6069336
      142.8208008,2.6069336 142.8208008,1.0351563 139.1640625,1.0351563 139.1640625,1.8208008
      135.5087891,1.8208008 135.5087891,11.25 131.8540039,11.25 131.8540039,9.6782227 128.1992188,9.6782227
      128.1992188,1.8208008 124.5429688,1.8208008 124.5429688,6.5361328 120.887207,6.5361328
      120.887207,12.8208008 117.230957,12.8208008 117.230957,10.4628906 113.5761719,10.4628906
      113.5761719,12.0341797 109.9199219,12.0341797 109.9199219,11.25 106.2651367,11.25 106.2651367,1.8208008
      102.6079102,1.8208008 102.6079102,3.3930664 98.953125,3.3930664 98.953125,11.25 95.296875,11.25
      95.296875,6.5361328 91.6401367,6.5361328 91.6401367,0.25 87.9848633,0.25 87.9848633,2.6069336
      84.3310547,2.6069336 84.3310547,1.0351563 80.6757813,1.0351563 80.6757813,1.8208008 77.019043,1.8208008
      77.019043,11.25 73.362793,11.25 73.362793,9.6782227 69.7080078,9.6782227 69.7080078,1.8208008
      66.0507813,1.8208008 66.0507813,6.5361328 62.3959961,6.5361328 62.3959961,12.8208008 58.7412109,12.8208008
      58.7412109,10.4628906 55.0849609,10.4628906 55.0849609,12.0341797 51.4291992,12.0341797 51.4291992,11.25
      47.7739258,11.25 47.7739258,1.8208008 44.1181641,1.8208008 44.1181641,3.3930664 40.4628906,3.3930664
      40.4628906,11.25 36.8071289,11.25 36.8071289,6.5361328 33.1518555,6.5361328 33.1518555,0.25 29.4960938,0.25
       29.4960938,2.6069336 25.8398438,2.6069336 25.8398438,1.0351563 22.1850586,1.0351563 22.1850586,1.8208008
      18.5288086,1.8208008 18.5288086,11.25 14.8720703,11.25 14.8720703,9.6782227 11.2158203,9.6782227
      11.2158203,1.8208008 7.5610352,1.8208008 7.5610352,6.5361328 3.9057617,6.5361328 3.9057617,12.8208008
      0,12.8208008"/>
```

Graphical Set 3: Participant A (Granted Points)

```
<line id="XMLID_6_" class="st0" x1="3.7169952" y1="9.7649841" x2="0" y2="9.7649841"/>
<line id="XMLID_5_" class="st0" x1="3.7169952" y1="3.4229736" x2="0" y2="3.4229736"/>
<polyline id="XMLID_4_" class="st0" points="318.5429688,16.1069336 300.0131836,16.1069336 300.0131836,12.9350586
      296.3579102,12.9350586 296.3579102,16.1069336 292.7041016,16.1069336 292.7041016,12.1411133
      289.0478516,12.1411133 289.0478516,11.3510742 285.3920898,11.3510742 285.3920898,1.8369141
      281.7358398,1.8369141 281.7358398,3.4228516 278.0810547,3.4228516 278.0810547,16.1069336
      274.4248047,16.1069336 274.4248047,6.5917969 270.7680664,6.5917969 270.7680664,0.25 267.1118164,0.25
      267.1118164,2.6279297 263.4570313,2.6279297 263.4570313,1.0429688 259.8017578,1.0429688
      259.8017578,1.8369141 256.1469727,1.8369141 256.1469727,16.1069336 252.4912109,16.1069336
      252.4912109,9.7651367 248.8349609,9.7651367 248.8349609,1.8369141 245.1791992,1.8369141
      245.1791992,6.5917969 241.5239258,6.5917969 241.5239258,12.9350586 237.8691406,12.9350586
      237.8691406,10.5571289 234.2138672,10.5571289 234.2138672,12.1411133 230.5571289,12.1411133
      230.5571289,11.3510742 226.9018555,11.3510742 226.9018555,16.1069336 223.2470703,16.1069336
      223.2470703,3.4228516 219.5917969,3.4228516 219.5917969,11.3510742 215.934082,11.3510742
      215.934082,6.5917969 212.277832,6.5917969 212.277832,0.25 208.6240234,0.25 208.6240234,2.6279297
      204.9677734,2.6279297 204.9677734,1.0429688 201.3110352,1.0429688 201.3110352,1.8369141
      197.6547852,1.8369141 197.6547852,11.3510742 194,11.3510742 194,9.7651367 190.34375,9.7651367
      190.34375,1.8369141 186.6879883,1.8369141 186.6879883,6.5917969 183.0341797,6.5917969
      183.0341797,12.9350586 179.3779297,12.9350586 179.3779297,10.5571289 175.7211914,10.5571289
      175.7211914,12.1411133 172.065918,12.1411133 172.065918,11.3510742 168.4101563,11.3510742
      168.4101563,1.8369141 164.7539063,1.8369141 164.7539063,16.1069336 161.0981445,16.1069336
      161.0981445,11.3510742 157.4418945,11.3510742 157.4418945,6.5917969 153.7890625,6.5917969 153.7890625,0.25
      150.1318359,0.25 150.1318359,2.6279297 146.4760742,2.6279297 146.4760742,1.0429688 142.8208008,1.0429688
      142.8208008,1.8369141 139.1640625,1.8369141 139.1640625,16.1069336 117.230957,16.1069336
      117.230957,12.1411133 113.5761719,12.1411133 113.5761719,11.3510742 109.9199219,11.3510742
      109.9199219,1.8369141 106.2651367,1.8369141 106.2651367,3.4228516 102.6079102,3.4228516
      102.6079102,11.3510742 98.953125,11.3510742 98.953125,6.5917969 95.296875,6.5917969 95.296875,0.25
      91.6401367,0.25 91.6401367,2.6279297 87.9848633,2.6279297 87.9848633,1.0429688 84.3310547,1.0429688
      84.3310547,1.8369141 80.6757813,1.8369141 80.6757813,11.3510742 77.019043,11.3510742 77.019043,9.7651367
      73.362793,9.7651367 73.362793,1.8369141 69.7080078,1.8369141 69.7080078,6.5917969 66.0507813,6.5917969
      66.0507813,12.9350586 62.3959961,12.9350586 62.3959961,10.5571289 58.7412109,10.5571289
      58.7412109,12.1411133 55.0849609,12.1411133 55.0849609,11.3510742 51.4291992,11.3510742
      51.4291992,1.8369141 47.7739258,1.8369141 47.7739258,3.4228516 44.1181641,3.4228516 44.1181641,11.3510742
      40.4628906,11.3510742 40.4628906,6.5917969 36.8071289,6.5917969 36.8071289,0.25 33.1518555,0.25
      33.1518555,2.6279297 29.4960938,2.6279297 29.4960938,1.0429688 25.8398438,1.0429688 25.8398438,1.8369141
      22.1850586,1.8369141 22.1850586,11.3510742 18.5288086,11.3510742 18.5288086,9.7651367 14.8720703,9.7651367
      14.8720703,1.8369141 11.2158203,1.8369141 11.2158203,6.5917969 7.5610352,6.5917969 7.5610352,12.9350586
      3.9057617,12.9350586 3.9057617,10.5571289 0,10.5571289"/>
```

Graphical Set 4: Participant A (Predicted Busyness)

```
<line id="XMLID_2_" class="st0" x1="3.7169952" y1="14.9170227" x2="0" y2="14.9170227"/>
<line id="XMLID_5_" class="st0" x1="3.7169952" y1="0.25" x2="0" y2="0.25"/>
<polyline id="XMLID_6_" class="st0" points="318.5429688,3.184082 289.0478516,3.184082 289.0478516,6.1171875
      274.4248047,6.1171875 274.4248047,9.0507813 256.1469727,9.0507813 256.1469727,3.184082 237.8691406,3.184082
       237.8691406,6.1171875 234.2138672,6.1171875 234.2138672,9.0507813 230.5571289,9.0507813
      230.5571289,6.1171875 226.9018555,6.1171875 226.9018555,9.0507813 223.2470703,9.0507813
```

```
      223.2470703,6.1171875 208.6240234,6.1171875 208.6240234,9.0507813 194,9.0507813 194,6.1171875
      183.0341797,6.1171875 183.0341797,9.0507813 179.3779297,9.0507813 179.3779297,6.1171875
      175.7211914,6.1171875 175.7211914,3.184082 172.065918,3.184082 172.065918,6.1171875 161.0981445,6.1171875
      161.0981445,9.0507813 150.1318359,9.0507813 150.1318359,6.1171875 146.4760742,6.1171875
      146.4760742,3.184082 139.1640625,3.184082 139.1640625,9.0507813 135.5087891,9.0507813 135.5087891,6.1171875
       131.8540039,6.1171875 131.8540039,9.0507813 128.1992188,9.0507813 128.1992188,3.184082
      109.9199219,3.184082 109.9199219,6.1171875 106.2651367,6.1171875 106.2651367,9.0507813
      102.6079102,9.0507813 102.6079102,3.184082 95.296875,3.184082 95.296875,6.1171875 87.9848633,6.1171875
      87.9848633,3.184082 84.3310547,3.184082 84.3310547,6.1171875 80.6757813,6.1171875 80.6757813,11.9829102
      77.019043,11.9829102 77.019043,6.1171875 73.362793,6.1171875 73.362793,3.184082 66.0507813,3.184082
      66.0507813,6.1171875 55.0849609,6.1171875 55.0849609,9.0507813 51.4291992,9.0507813 51.4291992,11.9829102
      47.7739258,11.9829102 47.7739258,9.0507813 44.1181641,9.0507813 44.1181641,6.1171875 29.4960938,6.1171875
      29.4960938,9.0507813 25.8398438,9.0507813 25.8398438,6.1171875 22.1850586,6.1171875 22.1850586,9.0507813
      18.5288086,9.0507813 18.5288086,3.184082 14.8720703,3.184082 14.8720703,6.1171875 11.2158203,6.1171875
      11.2158203,3.184082 7.5610352,3.184082 7.5610352,9.0507813 3.9057617,9.0507813 3.9057617,6.1171875
      0,6.1171875"/>
```

Graphical Set 5: Participant A (Predicted Stress)

```
<line id="XMLID_2_" class="st0" x1="3.7169952" y1="15.0470276" x2="0" y2="15.0470276"/>
<line id="XMLID_5_" class="st0" x1="3.7169952" y1="0.25" x2="0" y2="0.25"/>
<polyline id="XMLID_6_" class="st0" points="318.5429688,6.1689453 310.9819336,6.1689453 310.9819336,3.2089844
      300.0131836,3.2089844 300.0131836,0.25 292.7041016,0.25 292.7041016,3.2089844 289.0478516,3.2089844
      289.0478516,6.1689453 274.4248047,6.1689453 274.4248047,9.1269531 256.1469727,9.1269531
      256.1469727,3.2089844 237.8691406,3.2089844 237.8691406,6.1689453 234.2138672,6.1689453
      234.2138672,9.1269531 230.5571289,9.1269531 230.5571289,6.1689453 219.5917969,6.1689453
      219.5917969,3.2089844 212.277832,3.2089844 212.277832,6.1689453 208.6240234,6.1689453 208.6240234,9.1269531
       204.9677734,9.1269531 204.9677734,6.1689453 183.0341797,6.1689453 183.0341797,9.1269531
      179.3779297,9.1269531 179.3779297,3.2089844 172.065918,3.2089844 172.065918,6.1689453 157.4418945,6.1689453
       157.4418945,9.1269531 153.7890625,9.1269531 153.7890625,6.1689453 135.5087891,6.1689453
      135.5087891,9.1269531 131.8540039,9.1269531 131.8540039,12.0878906 128.1992188,12.0878906
      128.1992188,3.2089844 124.5429688,3.2089844 124.5429688,6.1689453 120.887207,6.1689453 120.887207,3.2089844
       109.9199219,3.2089844 109.9199219,6.1689453 106.2651367,6.1689453 106.2651367,9.1269531
      102.6079102,9.1269531 102.6079102,3.2089844 98.953125,3.2089844 98.953125,6.1689453 91.6401367,6.1689453
      91.6401367,9.1269531 87.9848633,9.1269531 87.9848633,3.2089844 84.3310547,3.2089844 84.3310547,9.1269531
      77.019043,9.1269531 77.019043,3.2089844 66.0507813,3.2089844 66.0507813,6.1689453 62.3959961,6.1689453
      62.3959961,9.1269531 55.0849609,9.1269531 55.0849609,6.1689453 51.4291992,6.1689453 51.4291992,9.1269531
      47.7739258,9.1269531 47.7739258,6.1689453 36.8071289,6.1689453 36.8071289,9.1269531 25.8398438,9.1269531
      25.8398438,6.1689453 22.1850586,6.1689453 22.1850586,9.1269531 18.5288086,9.1269531 18.5288086,6.1689453
      7.5610352,6.1689453 7.5610352,9.1269531 0,9.1269531"/>
```

Graphical Set 6: Participant A (Actual Steps: Top Curve)

```
<line id="XMLID_6_" class="st0" x1="3.717701" y1="34.5572166" x2="0.0007058" y2="34.5572166"/>
<line id="XMLID_5_" class="st0" x1="3.717701" y1="24.3072186" x2="0.0007058" y2="24.3072186"/>
<polyline id="XMLID_2_" class="st0" points="318.293457,44.3364258 314.6376953,48.7729492 310.9819336,31.9199219
      310.9624023,31.9350586 307.3271484,46.5717773 303.6767578,38.6401367 303.652832,38.6616211
      300.0136719,55.34375 296.3481445,36.4580078 296.3134766,36.4697266 292.7045898,40.8886719
      289.0488281,42.2119141 285.3862305,34.9516602 281.7275391,17.9467773 281.6943359,17.9604492
      278.0673828,23.2587891 274.425293,43.4731445 270.7602539,33.2670898 270.7211914,33.2685547
      267.1123047,35.003418 263.4594727,22.5576172 263.4394531,22.5742188 259.8022461,39.0668945
      256.1445313,27.5117188 256.1157227,27.5288086 252.4921875,35.3989258 248.8505859,21.6533203
      248.777832,21.5732422 245.2314453,20.2900391 245.1694336,20.3256836 241.5244141,39.1035156
      237.8691406,37.8076172 234.2133789,21.6879883 234.1933594,21.7036133 230.5576172,36.3886719
      226.9233398,33.0810547 223.2358398,27.8334961 219.5854492,15.4287109 219.5527344,15.4433594
      215.9213867,21.0004883 212.2788086,37.5668945 208.6162109,31.4941406 204.9702148,12.0214844
      204.9560547,12.0341797 201.3115234,39.1445313 197.6455078,34.0844727 193.9873047,13.0297852
      193.9501953,13.0366211 190.3447266,16.2075195 186.6704102,20.6538086 183.034668,39.8637695
      179.3779297,37.4643555 179.3554688,37.3657227 175.7236328,21.2519531 175.7055664,21.2666016
      172.0668945,40.5908203 168.4023438,33.9316406 164.7597656,19.625 164.7446289,19.6391602
      161.0986328,53.3349609 157.4399414,18.296875 157.4257813,18.3081055 153.7890625,38.3051758
      150.1152344,34.1572266 146.4819336,23.7568359 146.4619141,23.7749023 142.8222656,43.0913086
      139.1567383,31.0385742 139.1220703,31.0522461 135.4931641,35.9443359 131.8549805,48.940918
      128.2080078,43.8911133 128.1806641,43.9165039 124.5429688,57.6962891 120.8964844,54.925293
      117.2324219,51.6069336 117.2124023,51.5390625 113.5766602,38.9926758 109.9130859,32.3974609
      106.2543945,13.4458008 106.2182617,13.456543 102.6088867,17.3959961 98.9609375,8.3535156
      98.940918,8.3730469 95.296875,33.6875 91.6162109,30.800293 87.9833984,24.137207 87.9487305,24.1542969
      84.3310547,29.421875 80.6640625,11.8652344 80.6254883,11.8725586 76.9931641,14.9858398
      73.3637695,21.7265625 69.722168,18.7758789 69.6972656,18.8051758 66.0512695,39.0683594
      62.3969727,37.2294922 58.7426758,21.1123047 58.7236328,21.1274414 55.0859375,38.2695313
      51.4291992,26.6894531 47.7680664,18.7158203 47.7294922,18.7255859 44.1186523,21.815918
      40.4775391,18.7963867 40.453125,18.8261719 36.8081055,40.3496094 33.152832,38.3310547 29.46875,23.3842773
      25.8505859,20.1723633 25.8217773,20.2001953 22.1855469,31.7182617 18.5258789,0.3286133 18.5112305,0.340332
      14.8730469,18.1191406 11.2626953,19.8710938 11.2011719,19.9418945 7.5688477,30.9521484 3.9067383,37.4384766
       0.2504883,23.9228516"/>
```

Graphical Set 7: Participant A (Actual Steps: Bottom Curve)

```
<line id="XMLID_6_" class="st0" x1="3.717701" y1="34.5572166" x2="0.0007058" y2="34.5572166"/>
<line id="XMLID_5_" class="st0" x1="3.717701" y1="24.3072186" x2="0.0007058" y2="24.3072186"/>
<polyline class="st0" points="0.25,23.921875 3.9121094,37.3881836 3.9433594,37.3725586 7.5610352,30.9760742
      11.2167969,19.8935547 14.8203125,18.1450195 14.8842773,18.0620117 18.5288086,0.25 22.1928711,31.6474609
      22.2119141,31.6347656 25.8413086,20.1386719 29.4960938,23.409668 29.5141602,23.4848633
      33.1401367,38.2807617 33.1982422,38.3569336 36.7939453,40.3217773 36.8183594,40.2895508
      40.4638672,18.7602539 44.0483398,21.7568359 44.1879883,21.7558594 47.7739258,18.6875 51.4233398,26.675293
      55.0830078,38.2041016 55.1040039,38.1865234 58.7421875,21.0400391 62.3842773,37.1757813
      62.4462891,37.2548828 66.0380859,39.0410156 66.0625,39.0092773 69.7089844,18.7397461 73.3583984,21.7011719
      73.3896484,21.6777344 77.0556641,14.9160156 80.6767578,11.8305664 84.3417969,29.3730469
      84.3740234,29.3598633 87.9863281,24.1010742 91.6787109,30.8823242 95.2827148,33.6523438
      95.3071289,33.6210938 98.9541016,8.2885742 102.6152344,17.3613281 102.6533203,17.3486328
      106.2666016,13.4038086 109.9370117,32.4736328 113.5629883,38.9677734 117.2573242,51.6303711
```

```
      120.8876953,54.9174805 124.5322266,57.6645508 124.559082,57.6362305 128.1992188,43.8466797
      131.847168,48.8959961 131.8745117,48.871582 135.5371094,35.8637695 139.1645508,30.9951172
      142.8178711,43.0244141 142.8383789,43.0068359 146.4770508,23.6918945 150.1611328,34.2436523
      153.777832,38.2617188 153.8022461,38.2348633 157.4428711,18.2158203 161.0991211,53.2480469
      161.109375,53.2387695 164.7543945,19.5493164 168.4248047,33.9916992 172.0576172,40.5366211
      172.081543,40.5136719 175.722168,21.1791992 179.3681641,37.4179688 179.4194336,37.4921875
      183.0219727,39.8330078 183.046875,39.8022461 186.6894531,20.5947266 190.3330078,16.2231445 194,12.9921875
      197.684082,34.1933594 201.2993164,39.0947266 201.3217773,39.0693359 204.96875,11.9428711
      208.6445313,31.5805664 212.2700195,37.5151367 212.2949219,37.4916992 215.9609375,20.9199219
      219.5927734,15.3823242 223.2685547,27.9106445 226.9033203,33.0620117 230.5473633,36.3525391
      230.5742188,36.3242188 234.2138672,21.6196289 237.8549805,37.7451172 237.9296875,37.8286133
      241.4741211,39.0854492 241.5356445,39.050293 245.1796875,20.2719727 248.8359375,21.59375
      252.4960938,35.34375 252.5249023,35.3271484 256.1474609,27.4580078 259.7993164,39.0029297
      259.8208008,38.984375 263.4580078,22.4916992 267.1225586,34.9804688 267.1616211,34.9799805
      270.769043,33.2456055 274.4199219,43.4082031 274.440918,43.3891602 278.1118164,23.1645508
      281.7368164,17.8984375 285.40625,35.0146484 289.0058594,42.1269531 289.1376953,42.1791992
      292.7270508,40.8623047 296.3588867,36.4135742 300.0151367,55.2714844 300.0336914,55.2568359
      303.6704102,38.5825195 307.3217773,46.5151367 307.3466797,46.4941406 310.9829102,31.8500977
      314.6479492,48.7290039 314.6826172,48.7172852 318.3095703,44.3061523"/>
```

Graphical Set 8: Participant B (Goals)

```
<line id="XMLID_6_" class="st0" x1="0" y1="5.9415894" x2="4.5059662" y2="5.9415894"/>
<line id="XMLID_5_" class="st0" x1="0" y1="2.1107788" x2="4.5059662" y2="2.1107788"/>
<polyline id="XMLID_4_" class="st1" points="0,1.2641602 5.1204071,1.2641602 5.1204071,0 10.2408218,0
      10.2408218,6.2595825 15.3612289,6.2595825 15.3612289,8.7170105 20.4816437,8.7170105 20.4816437,1.298645
      25.6020508,1.298645 25.6020508,7.5830994 30.7224655,7.5830994 30.7224655,7.6271667 35.8428802,7.6271667
      35.8428802,0.2528381 40.9632797,0.2528381 40.9632797,1.2297058 46.0836906,1.2297058 46.0836906,0.0095825
      51.204113,0.0095825 51.204113,6.2767944 56.3245201,6.2767944 56.3245201,8.6825562 61.4449272,8.6825562
      61.4449272,1.3503723 66.5653381,1.3503723 66.5653381,7.5314026 71.6857605,7.5314026 71.6857605,7.678894
      76.8061676,7.678894 76.8061676,0.2088013 81.9265747,0.2088013 81.9265747,1.2641602 87.0469818,1.2641602
      87.0469818,0 92.1674042,0 92.1674042,6.2595825 97.2878113,6.2595825 97.2878113,8.7170105
      102.4082184,8.7170105 102.4082184,1.298645 107.5286255,1.298645 107.5286255,7.5830994 112.6490479,7.5830994
       112.6490479,7.6271667 117.769455,7.6271667 117.769455,0.2528381 122.8898621,0.2528381
      122.8898621,1.2297058 128.0102844,1.2297058 128.0102844,0.0095825 133.1306763,0.0095825
      133.1306763,6.2767944 138.2510834,6.2767944 138.2510834,8.6825562 143.3715057,8.6825562
      143.3715057,1.3503723 148.4919128,1.3503723 148.4919128,7.5314026 153.6123199,7.5314026
      153.6123199,7.678894 158.7327423,7.678894 158.7327423,0.2088013 163.8531342,0.2088013 163.8531342,1.2641602
       168.9735718,1.2641602 168.9735718,0 174.0939789,0 174.0939789,6.2595825 179.2143707,6.2595825
      179.2143707,8.7170105 184.3348083,8.7170105 184.3348083,1.298645 189.4552002,1.298645 189.4552002,7.5830994
       194.575592,7.5830994 194.575592,7.6271667 199.6960144,7.6271667 199.6960144,0.2528381
      204.8164368,0.2528381 204.8164368,1.2297058 209.9368591,1.2297058 209.9368591,0.0095825
      215.057251,0.0095825 215.057251,6.2767944 220.1776428,6.2767944 220.1776428,8.6825562 225.2980957,8.6825562
       225.2980957,1.3503723 230.4184875,1.3503723 230.4184875,7.5314026 235.5388947,7.5314026
      235.5388947,7.678894 240.659317,7.678894 240.659317,0.2088013 245.7797089,0.2088013 245.7797089,1.2641602
      250.9001007,1.2641602 250.9001007,0 256.0205383,0 256.0205383,6.2595825 261.1409302,6.2595825
      261.1409302,8.7170105 266.261322,8.7170105 266.261322,1.298645 271.3817139,1.298645 271.3817139,7.5830994
      276.5021667,7.5830994 276.5021667,7.6271667 281.6225586,7.6271667 281.6225586,0.2528381
      286.7429504,0.2528381 286.7429504,1.2297058 291.8634033,1.2297058 291.8634033,0.0095825
      296.9837952,0.0095825 296.9837952,6.2767944 302.104187,6.2767944 302.104187,8.6825562 307.2246094,8.6825562
       307.2246094,1.3503723 312.3450623,1.3503723 312.3450623,7.5314026 317.4654236,7.5314026
      317.4654236,7.678894 322.5858459,7.678894 322.5858459,0.2088013 327.7062378,0.2088013 327.7062378,1.2641602
       332.8266907,1.2641602 332.8266907,0 337.9470825,0 337.9470825,6.2595825 343.0674744,6.2595825
      343.0674744,8.7170105 348.1879272,8.7170105 348.1879272,1.298645 353.3082886,1.298645 353.3082886,7.5830994
       358.4286804,7.5830994 358.4286804,7.6271667 363.5491333,7.6271667 363.5491333,0.2528381
      368.6695557,0.2528381 368.6695557,1.2297058 373.7899475,1.2297058 373.7899475,0.0095825
      378.9103394,0.0095825 378.9103394,6.2767944 384.0307312,6.2767944 384.0307312,8.6825562
      389.1511841,8.6825562 389.1511841,1.3503723 394.2716064,1.3503723 394.2716064,7.5314026
      399.3919983,7.5314026 399.3919983,7.678894 404.5124207,7.678894 404.5124207,0.2088013 409.6328125,0.2088013
       409.6328125,1.2641602 414.7532043,1.2641602 414.7532043,0 419.8736572,0 419.8736572,6.2595825
      424.9940796,6.2595825 424.9940796,8.7170105 430.1144714,8.7170105 430.1144714,1.298645 435.2348633,1.298645
       435.2348633,7.5830994 440.3552246,7.5830994 440.3552246,7.6271667 445.475708,7.6271667
      445.475708,0.2528381"/>
```

Graphical Set 9: Participant B (Expected Points)

```
<line id="XMLID_5_" class="st0" x1="0" y1="8.3816223" x2="4.5059662" y2="8.3816223"/>
<line id="XMLID_3_" class="st0" x1="0" y1="1.6763306" x2="4.5059662" y2="1.6763306"/>
<polyline id="XMLID_2_" class="st1" points="0,5.4480591 5.1204071,5.4480591 5.1204071,0.8381653
      10.2408218,0.8381653 10.2408218,2.9335632 15.3612289,2.9335632 15.3612289,6.7052917 20.4816437,6.7052917
      20.4816437,5.4480591 25.6020508,5.4480591 25.6020508,5.867157 30.7224655,5.867157 30.7224655,6.2862244
      35.8428802,6.2862244 35.8428802,1.6763306 40.9632797,1.6763306 40.9632797,1.2572327 46.0836906,1.2572327
      46.0836906,5.867157 51.204113,5.867157 51.204113,3.7717285 56.3245201,3.7717285 56.3245201,0 61.4449272,0
      61.4449272,1.2572327 66.5653381,1.2572327 66.5653381,0.8381653 71.6857605,0.8381653 71.6857605,0.4190979
      76.8061676,0.4190979 76.8061676,5.0289612 81.9265747,5.0289612 81.9265747,5.4480591 87.0469818,5.4480591
      87.0469818,0.8381653 92.1674042,0.8381653 92.1674042,2.9335632 97.2878113,2.9335632 97.2878113,6.7052917
      102.4082184,6.7052917 102.4082184,5.4480591 107.5286255,5.4480591 107.5286255,5.867157 112.6490479,5.867157
       112.6490479,6.2862244 117.769455,6.2862244 117.769455,1.6763306 122.8898621,1.6763306
      122.8898621,1.2572327 128.0102844,1.2572327 128.0102844,5.867157 133.1306763,5.867157 133.1306763,3.7717285
       138.2510834,3.7717285 138.2510834,0 143.3715057,0 143.3715057,1.2572327 148.4919128,1.2572327
      148.4919128,0.8381653 153.6123199,0.8381653 153.6123199,0.4190979 158.7327423,0.4190979
      158.7327423,5.0289612 163.8531342,5.0289612 163.8531342,5.4480591 168.9735718,5.4480591
      168.9735718,0.8381653 174.0939789,0.8381653 174.0939789,2.9335632 179.2143707,2.9335632
      179.2143707,6.7052917 184.3348083,6.7052917 184.3348083,5.4480591 189.4552002,5.4480591
      189.4552002,5.867157 194.575592,5.867157 194.575592,6.2862244 199.6960144,6.2862244 199.6960144,1.6763306
      204.8164368,1.6763306 204.8164368,1.2572327 209.9368591,1.2572327 209.9368591,5.867157 215.057251,5.867157
      215.057251,3.7717285 220.1776428,3.7717285 220.1776428,0 225.2980957,0 225.2980957,1.2572327
      230.4184875,1.2572327 230.4184875,0.8381653 235.5388947,0.8381653 235.5388947,0.4190979
      240.659317,0.4190979 240.659317,5.0289612 245.7797089,5.0289612 245.7797089,5.4480591 250.9001007,5.4480591
       250.9001007,0.8381653 256.0205383,0.8381653 256.0205383,2.9335632 261.1409302,2.9335632
      261.1409302,6.7052917 266.261322,6.7052917 266.261322,5.4480591 271.3817139,5.4480591 271.3817139,5.867157
      276.5021667,5.867157 276.5021667,6.2862244 281.6225586,6.2862244 281.6225586,1.6763306
      286.7429504,1.6763306 286.7429504,1.2572327 291.8634033,1.2572327 291.8634033,5.867157 296.9837952,5.867157
```

```
 296.9837952,3.7717285 302.104187,3.7717285 302.104187,0 307.2246094,0 307.2246094,1.2572327
      312.3450623,1.2572327 312.3450623,0.8381653 317.4654236,0.8381653 317.4654236,0.4190979
      322.5858459,0.4190979 322.5858459,5.0289612 327.7062378,5.0289612 327.7062378,5.4480591
      332.8266907,5.4480591 332.8266907,0.8381653 337.9470825,0.8381653 337.9470825,2.9335632
      343.0674744,2.9335632 343.0674744,6.7052917 348.1879272,6.7052917 348.1879272,5.4480591
      353.3082886,5.4480591 353.3082886,5.867157 358.4286804,5.867157 358.4286804,6.2862244 363.5491333,6.2862244
       363.5491333,1.6763306 368.6695557,1.6763306 368.6695557,1.2572327 373.7899475,1.2572327
      373.7899475,5.867157 378.9103394,5.867157 378.9103394,3.7717285 384.0307312,3.7717285 384.0307312,0
      389.1511841,0 389.1511841,1.2572327 394.2716064,1.2572327 394.2716064,0.8381653 399.3919983,0.8381653
      399.3919983,0.4190979 404.5124207,0.4190979 404.5124207,5.0289612 409.6328125,5.0289612
      409.6328125,5.4480591 414.7532043,5.4480591 414.7532043,0.8381653 419.8736572,0.8381653
      419.8736572,2.9335632 424.9940796,2.9335632 424.9940796,6.7052917 430.1144714,6.7052917
      430.1144714,5.4480591 435.2348633,5.4480591 435.2348633,5.867157 440.3552246,5.867157 440.3552246,6.2862244
       445.475708,6.2862244 445.475708,1.6763306"/>
```

Graphical Set 10: Participant B (Granted Points)

```
<line id="XMLID_4_" class="st0" x1="0" y1="8.3816223" x2="4.5059662" y2="8.3816223"/>
<line id="XMLID_3_" class="st0" x1="0" y1="1.6763306" x2="4.5059662" y2="1.6763306"/>
<polyline id="XMLID_2_" class="st1" points="0,5.0289612 5.1204071,5.0289612 5.1204071,5.4480438
      10.2408218,5.4480438 10.2408218,0.83815 15.3612289,0.83815 15.3612289,2.9335632 20.4816437,2.9335632
      20.4816437,6.7052917 25.6020508,6.7052917 25.6020508,5.4480438 30.7224655,5.4480438 30.7224655,5.8671417
      35.8428802,5.8671417 35.8428802,6.2862244 40.9632797,6.2862244 40.9632797,1.6763306 46.0836906,1.6763306
      46.0836906,1.2572327 51.204113,1.2572327 51.204113,5.8671417 56.3245201,5.8671417 56.3245201,3.7717133
      61.4449272,3.7717133 61.4449272,0 66.5653381,0 66.5653381,1.2572327 71.6857605,1.2572327 71.6857605,0.83815
       76.8061676,0.83815 76.8061676,0.4190826 81.9265747,0.4190826 81.9265747,5.0289612 87.0469818,5.0289612
      87.0469818,5.4480438 92.1674042,5.4480438 92.1674042,0.83815 97.2878113,0.83815 97.2878113,2.9335632
      102.4082184,2.9335632 102.4082184,6.7052917 107.5286255,6.7052917 107.5286255,5.4480438
      112.6490479,5.4480438 112.6490479,5.8671417 117.769455,5.8671417 117.769455,6.2862244 122.8898621,6.2862244
       122.8898621,1.6763306 128.0102844,1.6763306 128.0102844,1.2572327 133.1306763,1.2572327
      133.1306763,5.8671417 138.2510834,5.8671417 138.2510834,3.7717133 143.3715057,3.7717133 143.3715057,0
      148.4919128,0 148.4919128,1.2572327 153.6123199,1.2572327 153.6123199,0.83815 158.7327423,0.83815
      158.7327423,0.4190826 163.8531342,0.4190826 163.8531342,5.0289612 168.9735718,5.0289612
      168.9735718,5.4480438 174.0939789,5.4480438 174.0939789,0.83815 179.2143707,0.83815 179.2143707,2.9335632
      184.3348083,2.9335632 184.3348083,6.7052917 189.4552002,6.7052917 189.4552002,5.4480438
      194.575592,5.4480438 194.575592,5.8671417 199.6960144,5.8671417 199.6960144,6.2862244 204.8164368,6.2862244
       204.8164368,1.6763306 209.9368591,1.6763306 209.9368591,1.2572327 215.057251,1.2572327
      215.057251,5.8671417 220.1776428,5.8671417 220.1776428,3.7717133 225.2980957,3.7717133 225.2980957,0
      230.4184875,0 230.4184875,1.2572327 235.5388947,1.2572327 235.5388947,0.83815 240.659317,0.83815
      240.659317,0.4190826 245.7797089,0.4190826 245.7797089,5.0289612 250.9001007,5.0289612
      250.9001007,5.4480438 256.0205383,5.4480438 256.0205383,0.83815 261.1409302,0.83815 261.1409302,2.9335632
      266.261322,2.9335632 266.261322,6.7052917 271.3817139,6.7052917 271.3817139,5.4480438 276.5021667,5.4480438
       276.5021667,5.8671417 281.6225586,5.8671417 281.6225586,6.2862244 286.7429504,6.2862244
      286.7429504,1.6763306 291.8634033,1.6763306 291.8634033,1.2572327 296.9837952,1.2572327
      296.9837952,8.3816223 302.104187,8.3816223 302.104187,3.7717133 307.2246094,3.7717133 307.2246094,0
      312.3450623,0 312.3450623,1.2572327 317.4654236,1.2572327 317.4654236,0.83815 322.5858459,0.83815
      322.5858459,0.4190826 327.7062378,0.4190826 327.7062378,5.0289612 332.8266907,5.0289612
      332.8266907,5.4480438 337.9470825,5.4480438 337.9470825,0.83815 343.0674744,0.83815 343.0674744,2.9335632
      348.1879272,2.9335632 348.1879272,6.7052917 353.3082886,6.7052917 353.3082886,5.4480438
      358.4286804,5.4480438 358.4286804,5.8671417 363.5491333,5.8671417 363.5491333,6.2862244
      368.6695557,6.2862244 368.6695557,1.6763306 373.7899475,1.6763306 373.7899475,1.2572327
      378.9103394,1.2572327 378.9103394,8.3816223 384.0307312,8.3816223 384.0307312,3.7717133
      389.1511841,3.7717133 389.1511841,0 394.2716064,0 394.2716064,1.2572327 399.3919983,1.2572327
      399.3919983,0.83815 404.5124207,0.83815 404.5124207,0.4190826 409.6328125,0.4190826 409.6328125,5.0289612
      414.7532043,5.0289612 414.7532043,8.3816223 419.8736572,8.3816223 419.8736572,0.83815 424.9940796,0.83815
      424.9940796,2.9335632 430.1144714,2.9335632 430.1144714,6.7052917 435.2348633,6.7052917
      435.2348633,5.4480438 440.3552246,5.4480438 440.3552246,5.8671417 445.475708,5.8671417
      445.475708,6.2862244"/>
```

Graphical Set 11: Participant B (Predicted Busyness)

```
<line id="XMLID_5_" class="st0" x1="0" y1="7.5704956" x2="4.5059662" y2="7.5704956"/>
<line id="XMLID_4_" class="st0" x1="0" y1="0" x2="4.5059662" y2="0"/>
<polyline id="XMLID_3_" class="st1" points="0,4.5422974 5.1204071,4.5422974 10.2408218,4.5422974
      15.3612289,4.5422974 15.3612289,3.0281982 20.4816437,3.0281982 25.6020508,3.0281982 30.7224655,3.0281982
      30.7224655,4.5422974 35.8428802,4.5422974 40.9632797,4.5422974 46.0836906,4.5422974 51.204113,4.5422974
      51.204113,3.0281982 56.3245201,3.0281982 61.4449272,3.0281982 66.5653381,3.0281982 66.5653381,4.5422974
      71.6857605,4.5422974 76.8061676,4.5422974 81.9265747,4.5422974 81.9265747,3.0281982 87.0469818,3.0281982
      87.0469818,4.5422974 92.1674042,4.5422974 97.2878113,4.5422974 97.2878113,3.0281982 102.4082184,3.0281982
      107.5286255,3.0281982 107.5286255,6.0563965 112.6490479,6.0563965 112.6490479,4.5422974
      117.769455,4.5422974 122.8898621,4.5422974 122.8898621,3.0281982 128.0102844,3.0281982
      133.1306763,3.0281982 133.1306763,4.5422974 138.2510834,4.5422974 143.3715057,4.5422974
      148.4919128,4.5422974 153.6123199,4.5422974 158.7327423,4.5422974 158.7327423,3.0281982
      163.8531342,3.0281982 163.8531342,4.5422974 168.9735718,4.5422974 174.0939789,4.5422974
      179.2143707,4.5422974 184.3348083,4.5422974 189.4552002,4.5422974 189.4552002,3.0281982
      194.575592,3.0281982 194.575592,4.5422974 199.6960144,4.5422974 204.8164368,4.5422974 209.9368591,4.5422974
       215.057251,4.5422974 220.1776428,4.5422974 225.2980957,4.5422974 230.4184875,4.5422974
      235.5388947,4.5422974 240.659317,4.5422974 245.7797089,4.5422974 250.9001007,4.5422974
      256.0205383,4.5422974 256.0205383,6.0563965 261.1409302,6.0563965 261.1409302,4.5422974
      266.261322,4.5422974 271.3817139,4.5422974 276.5021667,4.5422974 281.6225586,4.5422974
      286.7429504,4.5422974 291.8634033,4.5422974 291.8634033,6.0563965 296.9837952,6.0563965
      296.9837952,4.5422974 302.104187,4.5422974 307.2246094,4.5422974 312.3450623,4.5422974
      317.4654236,4.5422974 322.5858459,4.5422974 327.7062378,4.5422974 332.8266907,4.5422974
      337.9470825,4.5422974 343.0674744,4.5422974 348.1879272,4.5422974 353.3082886,4.5422974
      358.4286804,4.5422974 363.5491333,4.5422974 363.5491333,6.0563965 368.6695557,6.0563965
      368.6695557,4.5422974 373.7899475,4.5422974 378.9103394,4.5422974 384.0307312,4.5422974
      389.1511841,4.5422974 394.2716064,4.5422974 399.3919983,4.5422974 399.3919983,6.0563965
      404.5124207,6.0563965 404.5124207,4.5422974 409.6328125,4.5422974 414.7532043,4.5422974
      419.8736572,4.5422974 424.9940796,4.5422974 430.1144714,4.5422974 435.2348633,4.5422974
      435.2348633,6.0563965 440.3552246,6.0563965 440.3552246,4.5422974 445.475708,4.5422974"/>
```

Graphical Set 12: Participant B (Predicted Stress)

```
<line id="XMLID_5_" class="st0" x1="0" y1="7.8228607" x2="4.5059662" y2="7.8228607"/>
<line id="XMLID_4_" class="st0" x1="0" y1="0" x2="4.5059662" y2="0"/>
<polyline id="XMLID_3_" class="st1" points="0,4.6937103 5.1204071,4.6937103 10.2408218,4.6937103
      15.3612289,4.6937103 20.4816437,4.6937103 25.6020508,4.6937103 30.7224655,4.6937103 35.8428802,4.6937103
      40.9632797,4.6937103 46.0836906,4.6937103 51.204113,4.6937103 56.3245201,4.6937103 61.4449272,4.6937103
      66.5653381,4.6937103 71.6857605,4.6937103 76.8061676,4.6937103 81.9265747,4.6937103 87.0469818,4.6937103
      92.1674042,4.6937103 97.2878113,4.6937103 102.4082184,4.6937103 102.4082184,3.1291504 107.5286255,3.1291504
       107.5286255,6.2583008 112.6490479,6.2583008 112.6490479,4.6937103 117.769455,4.6937103
      122.8898621,4.6937103 128.0102844,4.6937103 133.1306763,4.6937103 138.2510834,4.6937103
      143.3715057,4.6937103 148.4919128,4.6937103 148.4919128,6.2583008 153.6123199,6.2583008
      153.6123199,4.6937103 158.7327423,4.6937103 163.8531342,4.6937103 168.9735718,4.6937103
      174.0939789,4.6937103 174.0939789,6.2583008 179.2143707,6.2583008 184.3348083,6.2583008
      189.4552002,6.2583008 189.4552002,4.6937103 194.575592,4.6937103 199.6960144,4.6937103
      204.8164368,4.6937103 209.9368591,4.6937103 215.057251,4.6937103 220.1776428,4.6937103
      225.2980957,4.6937103 230.4184875,4.6937103 235.5388947,4.6937103 240.659317,4.6937103
      245.7797089,4.6937103 250.9001007,4.6937103 256.0205383,4.6937103 256.0205383,6.2583008
      261.1409302,6.2583008 261.1409302,4.6937103 266.261322,4.6937103 271.3817139,4.6937103
      276.5021667,4.6937103 281.6225586,4.6937103 286.7429504,4.6937103 291.8634033,4.6937103
      291.8634033,6.2583008 296.9837952,6.2583008 296.9837952,4.6937103 302.104187,4.6937103
      307.2246094,4.6937103 312.3450623,4.6937103 317.4654236,4.6937103 322.5858459,4.6937103
      327.7062378,4.6937103 332.8266907,4.6937103 337.9470825,4.6937103 343.0674744,4.6937103
      348.1879272,4.6937103 353.3082886,4.6937103 358.4286804,4.6937103 363.5491333,4.6937103
      363.5491333,6.2583008 368.6695557,6.2583008 368.6695557,4.6937103 373.7899475,4.6937103
      378.9103394,4.6937103 384.0307312,4.6937103 389.1511841,4.6937103 394.2716064,4.6937103
      399.3919983,4.6937103 399.3919983,6.2583008 404.5124207,6.2583008 404.5124207,4.6937103
      409.6328125,4.6937103 414.7532043,4.6937103 419.8736572,4.6937103 424.9940796,4.6937103
      430.1144714,4.6937103 435.2348633,4.6937103 435.2348633,6.2583008 440.3552246,6.2583008
      440.3552246,4.6937103 445.475708,4.6937103"/>
```

Graphical Set 13: Participant B (Predicted Typical)

```
<line id="XMLID_5_" class="st0" x1="0" y1="7.8228302" x2="4.5059662" y2="7.8228302"/>
<line id="XMLID_4_" class="st0" x1="0" y1="0" x2="4.5059662" y2="0"/>
<polyline id="XMLID_3_" class="st1" points="0,3.1291199 5.1204071,3.1291199 10.2408218,3.1291199
      15.3612289,3.1291199 20.4816437,3.1291199 25.6020508,3.1291199 30.7224655,3.1291199 35.8428802,3.1291199
      40.9632797,3.1291199 46.0836906,3.1291199 51.204113,3.1291199 56.3245201,3.1291199 61.4449272,3.1291199
      66.5653381,3.1291199 71.6857605,3.1291199 71.6857605,4.6937103 76.8061676,4.6937103 76.8061676,1.5645599
      81.9265747,1.5645599 81.9265747,3.1291199 87.0469818,3.1291199 92.1674042,3.1291199 92.1674042,4.6937103
      97.2878113,4.6937103 97.2878113,3.1291199 102.4082184,3.1291199 102.4082184,4.6937103 107.5286255,4.6937103
       107.5286255,6.2582703 112.6490479,6.2582703 112.6490479,3.1291199 117.769455,3.1291199
      117.769455,4.6937103 122.8898621,4.6937103 122.8898621,3.1291199 128.0102844,3.1291199
      133.1306763,3.1291199 138.2510834,3.1291199 138.2510834,4.6937103 143.3715057,4.6937103
      148.4919128,4.6937103 148.4919128,3.1291199 153.6123199,3.1291199 158.7327423,3.1291199
      163.8531342,3.1291199 168.9735718,3.1291199 174.0939789,3.1291199 179.2143707,3.1291199
      179.2143707,4.6937103 184.3348083,4.6937103 184.3348083,3.1291199 189.4552002,3.1291199
      194.575592,3.1291199 199.6960144,3.1291199 204.8164368,3.1291199 209.9368591,3.1291199 215.057251,3.1291199
       220.1776428,3.1291199 220.1776428,4.6937103 225.2980957,4.6937103 225.2980957,3.1291199
      230.4184875,3.1291199 230.4184875,4.6937103 235.5388947,4.6937103 235.5388947,3.1291199
      240.659317,3.1291199 245.7797089,3.1291199 245.7797089,4.6937103 250.9001007,4.6937103
      250.9001007,3.1291199 256.0205383,3.1291199 261.1409302,3.1291199 266.261322,3.1291199
      271.3817139,3.1291199 271.3817139,4.6937103 276.5021667,4.6937103 281.6225586,4.6937103
      286.7429504,4.6937103 291.8634033,4.6937103 291.8634033,3.1291199 296.9837952,3.1291199
      302.104187,3.1291199 307.2246094,3.1291199 312.3450623,3.1291199 317.4654236,3.1291199
      322.5858459,3.1291199 327.7062378,3.1291199 332.8266907,3.1291199 337.9470825,3.1291199
      343.0674744,3.1291199 348.1879272,3.1291199 353.3082886,3.1291199 358.4286804,3.1291199
      363.5491333,3.1291199 368.6695557,3.1291199 373.7899475,3.1291199 378.9103394,3.1291199
      378.9103394,4.6937103 384.0307312,4.6937103 389.1511841,4.6937103 389.1511841,3.1291199
      394.2716064,3.1291199 399.3919983,3.1291199 404.5124207,3.1291199 409.6328125,3.1291199
      409.6328125,4.6937103 414.7532043,4.6937103 419.8736572,4.6937103 424.9940796,4.6937103
      430.1144714,4.6937103 430.1144714,3.1291199 435.2348633,3.1291199 435.2348633,1.5645599
      440.3552246,1.5645599 440.3552246,3.1291199 445.475708,3.1291199"/>
```

Graphical Set 14: Participant B (Weekday - Weekend)

```
<line id="XMLID_4_" class="st0" x1="0" y1="9.3874207" x2="4.5059662" y2="9.3874207"/>
<line id="XMLID_3_" class="st0" x1="0" y1="0" x2="4.5059662" y2="0"/>
<polyline id="XMLID_2_" class="st1" points="0,0 5.1204071,0 10.2408218,0 10.2408218,9.3874207
      15.3612289,9.3874207 20.4816437,9.3874207 25.6020508,9.3874207 30.7224655,9.3874207 35.8428802,9.3874207
      35.8428802,0 40.9632797,0 46.0836906,0 46.0836906,9.3874207 51.204113,9.3874207 56.3245201,9.3874207
      61.4449272,9.3874207 66.5653381,9.3874207 71.6857605,9.3874207 71.6857605,0 76.8061676,0 81.9265747,0
      81.9265747,9.3874207 87.0469818,9.3874207 92.1674042,9.3874207 97.2878113,9.3874207 102.4082184,9.3874207
      107.5286255,9.3874207 107.5286255,0 112.6490479,0 117.769455,0 117.769455,9.3874207 122.8898621,9.3874207
      128.0102844,9.3874207 133.1306763,9.3874207 138.2510834,9.3874207 143.3715057,9.3874207 143.3715057,0
      148.4919128,0 153.6123199,0 153.6123199,9.3874207 158.7327423,9.3874207 163.8531342,9.3874207
      168.9735718,9.3874207 174.0939789,9.3874207 179.2143707,9.3874207 179.2143707,0 184.3348083,0 189.4552002,0
       189.4552002,9.3874207 194.575592,9.3874207 199.6960144,9.3874207 204.8164368,9.3874207
      209.9368591,9.3874207 215.057251,9.3874207 215.057251,0 220.1776428,0 225.2980957,0 225.2980957,9.3874207
      230.4184875,9.3874207 235.5388947,9.3874207 240.659317,9.3874207 245.7797089,9.3874207
      250.9001007,9.3874207 250.9001007,0 256.0205383,0 261.1409302,0 261.1409302,9.3874207 266.261322,9.3874207
      271.3817139,9.3874207 276.5021667,9.3874207 281.6225586,9.3874207 286.7429504,9.3874207 286.7429504,0
      291.8634033,0 296.9837952,0 296.9837952,9.3874207 302.104187,9.3874207 307.2246094,9.3874207
      312.3450623,9.3874207 317.4654236,9.3874207 322.5858459,9.3874207 322.5858459,0 327.7062378,0 332.8266907,0
       332.8266907,9.3874207 337.9470825,9.3874207 343.0674744,9.3874207 348.1879272,9.3874207
      353.3082886,9.3874207 358.4286804,9.3874207 358.4286804,0 363.5491333,0 368.6695557,0 368.6695557,9.3874207
       373.7899475,9.3874207 378.9103394,9.3874207 384.0307312,9.3874207 389.1511841,9.3874207
      394.2716064,9.3874207 394.2716064,0 399.3919983,0 404.5124207,0 404.5124207,9.3874207 409.6328125,9.3874207
       414.7532043,9.3874207 419.8736572,9.3874207 424.9940796,9.3874207 430.1144714,9.3874207 430.1144714,0
      435.2348633,0 440.3552246,0 440.3552246,9.3874207 445.475708,9.3874207"/>
```

Graphical Set 15: Participant B (Actual Steps)


```
<line id="XMLID_3_" class="st1" x1="0" y1="38.0544968" x2="4.5059662" y2="38.0544968"/>
<line id="XMLID_2_" class="st1" x1="0" y1="25.1239777" x2="4.5059662" y2="25.1239777"/>
<polyline id="XMLID_4_" class="st0" points="0,15.1545715 5.1204071,19.6737823 10.2408218,10.3444214
      15.3612289,33.9878464 20.4816437,45.6511688 25.6020508,19.7061157 30.7224655,40.4531059
      35.8428802,43.1685028 40.9632797,13.4089203 46.0836906,20.5530396 51.204113,13.5188599
      56.3245201,36.5157547 61.4449272,42.9745331 66.5653381,21.1090469 71.6857605,40.7698975
      76.8061676,31.033226 81.9265747,17.0812073 87.0469818,19.6479111 92.1674042,17.0812073
      97.2878113,36.4446335 102.4082184,39.8259735 107.5286255,19.796608 112.6490479,39.4833069
      117.769455,40.3302536 122.8898621,18.2320099 128.0102844,20.1845245 133.1306763,16.745018
      138.2510834,38.2096558 143.3715057,46.7437897 148.4919128,19.4798126 153.6123199,42.3150978
      158.7327423,42.9034386 163.8531342,16.5445938 168.9735718,11.1654968 174.0939789,0 179.2143707,36.6450577
      184.3348083,42.1146736 189.4552002,14.792511 194.575592,40.517746 199.6960144,37.226944
      204.8164368,12.7236099 209.9368591,21.5034409 215.057251,12.7882767 220.1776428,32.8628922
      225.2980957,45.3408318 230.4184875,16.4346848 235.5388947,40.7828369 240.659317,42.3603439
      245.7797089,17.798851 250.9001007,18.723381 256.0205383,13.1438599 261.1409302,37.0071259
      266.261322,41.946579 271.3817139,22.1629028 276.5021667,41.1448975 281.6225586,42.056488
      286.7429504,17.4044571 291.8634033,19.7772141 296.9837952,51.1336937 302.104187,37.2786713
      307.2246094,46.2976952 312.3450623,22.0077286 317.4654236,42.0306129 322.5858459,35.5007095
      327.7062378,17.139389 332.8266907,19.3052673 337.9470825,16.3376846 343.0674744,38.151474
      348.1879272,46.4722672 353.3082886,21.6779938 358.4286804,36.2506866 363.5491333,39.8065796
      368.6695557,15.8011017 373.7899475,20.7728577 378.9103394,59.1376991 384.0307312,37.7764664
      389.1511841,46.8924904 394.2716064,19.2406006 399.3919983,34.7507553 404.5124207,43.8796921
      409.6328125,17.7859116 414.7532043,61.8660355 419.8736572,12.0900269 424.9940796,37.750618
      430.1144714,35.4942551 435.2348633,20.2427063 440.3552246,42.8517113 445.475708,43.0844688"/>
```


Table A.1: Key for Scaling to Reported Ticks from retrieved Graphical Sets 1-15 Used for Chapter 4 (Participants A & B) and Chapter 5 (Participant B)

Just Walk Participant	Graphical Set#	Tick Line ID	Reported Real Value	Tick Line ID	Reported Real Value
A	1	XMLID_7_	6000	XMLID_5_	8000
	2	XMLID_3_	200	XMLID_2_	400
	3	XMLID_6_	200	XMLID_5_	400
	4	XMLID_2_	0	XMLID_5_	5
	5	XMLID_2_	0	XMLID_5_	5
	6	XMLID_6_	6000	XMLID_5_	8000
	7	XMLID_6_	6000	XMLID_5_	8000
B	8	XMLID_6_	6000	XMLID_5_	8000
	9	XMLID_5_	0	XMLID_3_	400
	10	XMLID_4_	0	XMLID_3_	400
	11	XMLID_5_	0	XMLID_4_	5
	12	XMLID_5_	0	XMLID_4_	5
	13	XMLID_5_	0	XMLID_4_	5
	14	XMLID_4_	0	XMLID_3_	1
	15	XMLID_3_	6000	XMLID_2_	8000

A.3 Final Processed Sets, 'Accuracy' & Confidence

Since cited source plots draw on a *linear* scale, the equation $y = mx + c$ is used for the recovery of the final set from any initially established arbitrary scale. Graphical Sets 1-15 are used with the scaling information provided in Table A.1.

Following appropriate rounding per the facts in Section A.2 and exclusion of re-

dundant points (i.e., staircase plots, `polygon` object), the final recovered sets from the described high-precision digitization procedure are provided in Tables A.2-A.6. Finally, estimated accuracy information and confidence statements are presented in Table A.7.

Table A.2: Final Processed Set (1/3) of Published *Just Walk* Plot from Mercere, 2017 (Participant A) Used in Chapter 4: Intervention Days 0-40
(see Table A.7 for Estimated Accuracy and Confidence)

Day	*Goals* steps/day	*Expected Points* 100-500	*Granted Points* 0-500	*Predicted Busyness* 1-4	*Predicted Stress* 1-5	*Behavior* (Δ_k) steps/day
0	5193	100	175	3	2	8075 (2)
1	4818	300	100	2	2	5438 (14)
2	8474	450	300	4	3	6699 (6)
3	8060	200	450	3	3	8861 (15)
4	8608	150	200	4	3	9207 (17)
5	5815	450	150	2	2	12694 (19)
6	4317	475	450	3	3	6554 (18)
7	7961	425	475	2	2	8813 (13)
8	5189	500	425	3	2	8180 (21)
9	4813	300	500	3	2	5264 (16)
10	8492	150	300	3	3	4870 (13)
11	8038	400	150	3	3	9082 (14)
12	8634	450	400	2	3	8486 (12)
13	5789	150	450	1	2	9097 (8)
14	4339	125	150	2	3	7538 (3)
15	7948	175	125	3	2	5276 (17)
16	5193	100	175	3	2	8637 (18)
17	4818	300	100	3	3	5479 (17)
18	8474	450	300	4	4	5120 (13)
19	8060	200	450	4	4	9086 (14)
20	8608	150	200	3	4	8504 (11)
21	5815	450	150	1	2	9819 (15)
22	4317	475	450	3	2	10434 (9)
23	7961	425	475	4	4	7002 (13)
24	5189	500	425	3	2	8040 (12)
25	4813	300	500	3	3	6717 (17)
26	8492	150	300	4	3	6170 (14)
27	8038	400	150	4	4	11126 (17)
28	8634	450	400	2	2	9349 (10)
29	5789	150	450	3	3	10127 (11)
30	4339	125	150	4	4	6421 (16)
31	7948	175	125	4	4	5139 (6)
32	5193	100	0	4	4	2669 (19)
33	4818	300	0	4	3	2026 (3)
34	8474	450	0	4	4	1485 (12)
35	8060	200	0	2	1	4187 (15)
36	8608	150	0	3	2	3193 (14)
37	5815	450	0	2	3	5745 (17)
38	4317	475	450	4	3	6695 (13)
39	7961	425	475	4	3	4335 (18)
40	5189	500	425	3	3	8120 (18)

Table A.3: Final Processed Set (2/3) of Published *Just Walk* Plot from Mercere, 2017 (Participant A) Used in Chapter 4: Intervention Days 41-80
(see Table A.7 for Estimated Accuracy and Confidence)

Day	*Goals* steps/day	*Expected Points* 100-500	*Granted Points* 0-500	*Predicted Busyness* 1-4	*Predicted Stress* 1-5	*Behavior* steps/day
41	4813	300	500	2	3	6078 (18)
42	8492	150	300	2	2	5269 (15)
43	8038	400	150	2	3	9189 (19)
44	8634	450	0	3	3	2336 (19)
45	5789	150	450	3	3	8928 (19)
46	4339	125	150	3	3	6122 (13)
47	7948	175	125	4	4	4823 (16)
48	5193	100	175	3	4	8610 (18)
49	4818	300	100	2	2	5442 (25)
50	8474	450	300	3	3	4965 (13)
51	8060	200	450	3	3	8713 (13)
52	8608	150	200	3	3	9577 (4)
53	5815	450	150	2	3	10208 (9)
54	4317	475	450	2	3	6071 (23)
55	7961	425	475	2	3	5105 (16)
56	5189	500	425	2	2	10413 (19)
57	4813	300	500	3	3	6581 (18)
58	8492	150	300	3	4	5413 (16)
59	8038	400	150	3	4	8661 (16)
60	8634	450	400	3	3	9741 (13)
61	5789	150	0	2	3	7312 (16)
62	4339	125	150	3	3	6288 (4)
63	7948	175	125	2	2	5643 (14)
64	5193	100	175	3	3	8524 (17)
65	4818	300	100	4	4	5366 (17)
66	8474	450	300	4	4	5113 (12)
67	8060	200	450	4	4	8787 (12)
68	8608	150	200	4	4	8529 (17)
69	5815	450	0	4	4	5836 (15)
70	4317	475	450	2	2	7385 (15)
71	7961	425	475	2	2	5120 (17)
72	5189	500	425	2	2	8354 (17)
73	4813	300	500	2	2	5913 (6)
74	8492	150	300	2	2	6256 (5)
75	8038	400	0	3	3	4260 (17)
76	8634	450	400	3	3	8223 (19)
77	5789	150	450	3	3	9250 (13)
78	4339	125	150	3	3	5923 (14)
79	7948	175	125	4	4	4506 (17)
80	5193	100	0	4	5	4770 (6)

Table A.4: Final Processed Set (3/3) of Published *Just Walk* Plot from Mercere, 2017 (Participant A) Used in Chapter 4: Intervention Days 81-87
(see Table A.7 for Estimated Accuracy and Confidence)

Day	*Goals* steps/day	*Expected Points* 100-500	*Granted Points* 0-500	*Predicted Busyness* 1-4	*Predicted Stress* 1-5	*Behavior* steps/day
81	4818	300	100	4	5	5638 (12)
82	8474	450	0	4	4	1944 (18)
83	8060	200	0	4	4	5215 (16)
84	8608	150	0	4	4	3656 (16)
85	5815	450	0	4	3	6528 (18)
86	4317	475	0	4	3	3226 (12)
87	7961	425	0	4	3	4092 (7)

Table A.5: Final Processed Set (1/2) of Published *Just Walk* Plot from Freigoun *et al.*, 2017 (Participant B) Used in Chapters 4 and 5: Intervention Days 0-50 (see Table A.7 for Estimated Accuracy and Confidence)

Day	*Goals* steps/day	*Expected Points* 100-500	*Granted Points* 0-500	*Predicted Busyness* 1-4	*Predicted Stress* 1-5	*Predicted Typical* 1-4	*Weekday/ Weekend* {0, 1}	*Behavior* steps/day
0	8442	175	200	2	2	3	1	9542
1	9102	450	175	2	2	3	1	8843
2	5834	325	450	2	2	3	0	10286
3	4551	100	325	3	2	3	0	6629
4	8424	175	100	3	2	3	0	4825
5	5143	150	175	3	2	3	0	8838
6	5120	125	150	2	2	3	0	5629
7	8970	400	125	2	2	3	1	5209
8	8460	425	400	2	2	3	1	9812
9	9097	150	425	2	2	3	0	8707
10	5825	275	150	3	2	3	0	9795
11	4569	500	275	3	2	3	0	6238
12	8397	425	500	3	2	3	0	5239
13	5170	450	425	2	2	3	0	8621
14	5093	475	450	2	2	2	1	5580
15	8993	200	475	2	2	4	1	7086
16	8442	175	200	3	2	3	0	9244
17	9102	450	175	2	2	3	0	8847
18	5834	325	450	2	2	2	0	9244
19	4551	100	325	3	2	3	0	6249
20	8424	175	100	3	3	2	0	5726
21	5143	150	175	1	1	1	1	8824
22	5120	125	150	2	2	3	1	5779
23	8970	400	125	2	2	2	0	5648
24	8460	425	400	3	2	3	0	9066
25	9097	150	425	3	2	3	0	8764
26	5825	275	150	2	2	3	0	9296
27	4569	500	275	2	2	2	0	5976
28	8397	425	500	2	2	2	1	4656
29	5170	450	425	2	1	3	1	8873
30	5093	475	450	2	2	3	0	5341
31	8993	200	475	3	2	3	0	5250
32	8442	175	200	2	2	3	0	9327
33	9102	450	175	2	2	3	0	10159
34	5834	325	450	2	1	3	0	11886
35	4551	100	325	2	1	2	1	6218
36	8424	175	100	2	1	3	1	5372
37	5143	150	175	3	2	3	0	9598
38	5120	125	150	2	2	3	0	5619
39	8970	400	125	2	2	3	0	6128
40	8460	425	400	2	2	3	0	9918
41	9097	150	425	2	2	3	0	8560
42	5825	275	150	2	2	3	1	9908
43	4569	500	275	2	2	2	1	6803
44	8397	425	500	2	2	3	0	4873
45	5170	450	425	2	2	2	0	9344
46	5093	475	450	2	2	3	0	5578
47	8993	200	475	2	2	3	0	5334
48	8442	175	200	2	2	2	0	9133
49	9102	450	175	2	2	3	1	8990
50	5834	325	450	1	1	3	1	9853

Table A.6: Final Processed Set (2/2) of Published *Just Walk* Plot from Freigoun *et al.*, 2017 (Participant B) Used in Chapters 4 and 5: Intervention Days 51-87 (see Table A.7 for Estimated Accuracy and Confidence)

Day	*Goals* steps/day	*Expected Points* 100-500	*Granted Points* 0-500	*Predicted Busyness* 1-4	*Predicted Stress* 1-5	*Predicted Typical* 1-4	*Weekday/ Weekend* {0, 1}	*Behavior* steps/day
51	4551	100	325	2	2	3	0	6162
52	8424	175	100	2	2	3	0	5398
53	5143	150	175	2	2	2	0	8458
54	5120	125	150	2	2	2	0	5522
55	8970	400	125	2	2	2	0	5381
56	8460	425	400	2	2	2	1	9194
57	9097	150	425	1	1	3	1	8827
58	5825	275	0	2	2	3	0	3977
59	4569	500	275	2	2	3	0	6120
60	8397	425	500	2	2	3	0	4725
61	5170	450	425	2	2	3	0	8482
62	5093	475	450	2	2	3	0	5385
63	8993	200	475	2	2	3	1	6395
64	8442	175	200	2	2	3	1	9235
65	9102	450	175	2	2	3	0	8900
66	5834	325	450	2	2	3	0	9359
67	4551	100	325	2	2	3	0	5985
68	8424	175	100	2	2	3	0	4698
69	5143	150	175	2	2	3	0	8533
70	5120	125	150	2	2	3	1	6279
71	8970	400	125	1	1	3	1	5729
72	8460	425	400	2	2	3	0	9442
73	9097	150	425	2	2	3	0	8673
74	5825	275	0	2	2	2	0	2739
75	4569	500	275	2	2	2	0	6043
76	8397	425	500	2	2	3	0	4633
77	5170	450	425	2	2	3	1	8910
78	5093	475	450	1	1	3	1	6511
79	8993	200	475	2	2	3	0	5099
80	8442	175	200	2	2	2	0	9135
81	9102	450	0	2	2	2	0	2317
82	5834	325	450	2	2	2	0	10016
83	4551	100	325	2	2	2	0	6047
84	8424	175	100	2	2	3	1	6396
85	5143	150	175	1	1	4	1	8755
86	5120	125	150	2	2	3	0	5258
87	8970	400	125	2	2	3	0	5222

Table A.7: Presentation of Round Off Accuracy and Confidence in the Recovery of Published Graphical Data in 1-15 from Cited Sources ($N_A = N_B = 88$ measurements)

Just Walk Participant	Graphical Set #	Rounding Accuracy		Confidence	Notes*
		Min	Mean		
A	1	99.18%	99.99%	Exact Recovery	16-day cycle verified & enforced numerically.
	2	98.09%	99.97%	Exact Recovery	16-day cycle verified & enforced numerically.
	3	99.81%	99.95%	Exact Recovery	"IF-THEN" verified numerically & visually.
	4	99.98%	99.99%	Exact Recovery	Confirmed visually.
	5	99.98%	99.99%	Exact Recovery	Confirmed visually.
	6, 7	N/A**	N/A**	$\lvert\Delta_k\rvert < 25$ steps/day†	Both sets are for *Behavior*; see notes below.†
B	8	99.9997%	99.9999%	Exact Recovery	16-day cycle verified numerically.
	9	99.9992%	99.9998%	Exact Recovery	16-day cycle verified numerically.
	10	99.9995%	99.9997%	Exact Recovery	"IF-THEN" verified numerically.
	11	99.9999%	99.9999%	Exact Recovery	Confirmed visually.
	12	99.9992%	99.9997%	Exact Recovery	Confirmed visually.
	13	99.9996%	99.9997%	Exact Recovery	Confirmed visually.
	14	100%	100%	Exact Recovery	Confirmed visually.
	15	99.9999%	99.9999%	Exact Recovery‡	See notes below.‡

*All conclusions in this tabulation and below are drawn with the strict use of published *Just Walk* information outlined in Section A.2 (sufficient to the extent of given statements).

**In this case, information from both `polyline` objects in Sets 6 and 7 (resulting from 'cutting' an obtained `polygon` object) are used jointly; only confidence can be estimated. Unique data points (in time) from both `polyline` objects are selected with an accuracy of a single `polyline` (i.e., similar to accuracy in all other sets).

†*Estimated* from $\lceil \max \Delta_k \rceil$, where $\Delta_k \in \mathbb{R}^{88}$ is the vector containing *maximum* integer gap values (absolute) between both `polyline` objects from Sets 6 and 7. *Estimated* accuracy is $> 99.07\%$. Minimum gap is 0 steps/day; maximum is 25 steps/day.

‡"Exact" is reported since the recovery/digitization procedure produced a single `polyline` object provided in Set 15 (with identical properties to Sets 1-5, 8-14), rendering the *estimated* accuracy $> 99.9999\%$ and hence the "exactness" from knowing that steps/day must be a whole number.