Adaptive Learning and Control in Dynamic Environments

Paulina Weiss

Master of Science
School of Informatics
University of Edinburgh
2023

Abstract

To interact effectively with the world around us, we need to learn how to control complex dynamic systems. In this project, we investigate human goal-directed behaviour in dynamic control tasks. Specifically, we draw on a novel type of control environment based on "Ornstein–Uhlenbeck" (OU) networks which simulate causally related variables over time through Gauss–Markov processes. We are interested in the extent to which effective control relies on developing representations of the causal structure and dynamics of this task environment. Recent studies found that people are best described by a simple control strategy when interacting with continuous-time OU networks, suggesting they represent only the minimal amount of structure needed to achieve a control goal (Davis et al., 2018, 2020b). Here, we examine control behaviour in a more challenging control task featuring multidimensional sequential decision-making and discrete time intervals. While participants exhibit behavioural markers that are linked to simpler control strategies, a more complex environment elicits interventions best described by models with a more sophisticated representation of the task structure.

Keywords: dynamic control, causal learning, dynamic systems, computational modelling, cognitive modelling, intervention, reinforcement learning

Research Ethics Approval

This project obtained approval from the PPLS Research Ethics Committee of the School of Philosophy, Psychology and Language Sciences.
Ethics application number: 354-2223/2
Date when approval was obtained: 2023-06-30
The participants' information sheet and a consent form are included in the appendix.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Paulina Weiss*)

Acknowledgements

I would like to thank my supervisor Neil Bramley for his consistent support and valuable guidance throughout the project. I am also grateful to Victor Btesh and Franziska Kaltenberger for their feedback, inspiration and sense of community.

Table of Contents

Chapter 1

Introduction

Many tasks in our day-to-day lives require us to control or interact with complex dynamic systems whose structure is initially unknown, ambiguous or inherently uncertain to us. A farmer, for example, needs to remain flexible to accommodate daily changes in weather patterns, as well as unforeseen events such as disease outbreaks or market fluctuations. Consequently, farmers must learn which components of their environment lead to a healthy yield to make timely and effective interventions on their crop's growing conditions. For example, they may need to adjust the amount of water in a different way if a crop is not growing well due to lack of sunlight than if it is due to poor soil quality. At the same time, they must remain sensitive to the potential impact of water adjustments on soil degradation or even the risk of crop damage.

Goal-directed learning and complex problem-solving are crucial cognitive activities to carry out effective actions in both everyday and professional settings. What makes interacting with naturalistic environments so complex is that they typically involve a combination of rich, often non-linear, and random relations between different variables that may be continuously changing, either as a direct consequence of our actions, autonomously, or both (Simon, 1962; Waldrop, 1993; Holland, 2014). Therefore, an important question is how people accommodate these complexities when learning to control a new task.

Recent work in causal cognition has explored many of the complexities of goal-directed learning in real-world scenarios, isolating factors like stochasticity, interventions, time, and continuous variables, both individually (e.g., Bramley et al., 2017, 2015, 2018) and in combination (e.g., Davis et al., 2020a; Rehder et al., 2022). These studies show that people effectively construct causal representations of the task environment simply by observing or intervening in a task when they are explicitly asked to do so. Additionally, a substantial body of work has examined people´s ability to control various complex dynamic tasks (Osman, 2010). However, a majority of these studies do not formalise the properties of more complex and realistic dynamic systems and, as a result, they are unable to discern how control performance relies on learning the system structure (Kirlik, 2007). Thus, the extent to which structure representations, particularly causal ones, influence control performance in tasks that are not aimed at maximising our knowledge about their underlying environment remains unclear.

In this paper, we investigate goal-directed behaviour in a challenging control task that does not explicitly engage the participants in causal structure learning. We draw on a novel class of dynamic systems developed by Davis et al. (2018, 2020b), capturing some of the complexities of real-world dynamic environments while maintaining a principled approach that enables formal analysis of causal structure learning. The paper is structured as follows. First, we review relevant prior work on causal learning and dynamic control. Next, we introduce the new framework for modelling dynamic systems and detail three distinct computational accounts of causal structure learning of those systems. We then report on a challenging control task where participants interacted with the system. We find that participants exhibit some of the behavioural markers linked to simpler control strategies. However, their interventions were best described by models with a more sophisticated representation of the task structure.

Chapter 2

Background and Related Work

2.1 Causal Learning

A common paradigm for understanding the mind posits that we inherently seek to develop and rely on rich, abstract representations of our environment in order to interact with the physical and social world around us (Tenenbaum et al., 2011). In line with this paradigm, a significant body of work on causal learning has established the relevance of representing causal structure for accurately predicting outcomes, providing explanations, and making effective responses within uncertain environments (Sloman, 2005).

Causal learning has been predominantly studied within static environments, where people interact with deterministic causal structures between discrete variables. A typical task, for example, would ask participants to discover the effect of a headache-relief pill based on several patients who either take the pill or not, and observe if their headaches get better or not. In such scenarios, causal structure can be defined in terms of probabilistic contingencies, where a cause influences the probability of its effect (Hitchcock, 2018). This probabilistic understanding of causality has paved the way for the adoption of a well-established framework known as Causal Bayesian Networks (Pearl, 2009) which has enabled researchers to formally represent the causal structure between multiple variables as networks of probabilistic contingencies. Some studies, for example, demonstrated that people's causal structure judgments reflect the actual contingencies between the variables in the network when they observed or intervened in the network (Lagnado and Sloman, 2004; Griffiths and Tenenbaum, 2009; Holyoak and Cheng, 2011). Similarly, research on active learning established that people can intervene in the networks in ways that are effectively reducing their uncertainty about the network´s true structure (Bramley et al., 2015; Coenen et al., 2015).

Describing causal structure in terms of contingencies has played a pivotal role in establishing a principled analysis of causality in cognition. However, in many real-world dynamic systems, people interact with causal relations that unfold over time and rely on information that is autocorrelated, noisy or real-valued (Sloman and Lagnado, 2015). Static task environments where temporal information is either uninformative or omitted are not equipped to describe these complexities. Several studies have explored more complex dynamics, finding that people are capable of constructing causal representations in light of stochasticity (Bramley et al., 2017; Rothe et al., 2018), amongst real-valued variables (Pacer and Griffiths, 2011), and for dynamics that unfold in time (Bramley et al., 2014, 2018). Collectively, these investigations underscore our capacity to accommodate some of the complexities inherent to realistic dynamic systems.

Most recently, Davis et al. (2020a) have developed a new type of task environment which enables the study of multiple complexities of real-world causal learning at the same time. In both of their papers, they simulate the dynamics of a set of causally related, real-valued variables over time using Gauss–Markov processes. This framework, referred to as an "Ornstein–Uhlenbeck" (OU) network, is able to model complex non-linear phenomena such as feedback loops or oscillations. Crucially, it allows for a systematic investigation of the types of causal representations people construct when interacting with the task. The authors find that causal judgements were most accurately described by a 'local computations' heuristic which represents causal relationships between pairs of variables, rather than a normative model that represents entire causal structures.

While research on causal learning has been valuable in demonstrating that people are capable of constructing causal representations when their goal is to discover the structure underlying a task environment, many everyday control tasks are not aimed at maximising our knowledge. Here, we explore this aspect further by asking how and to what extent people build structure representations when they are engaged in a task other than discovering causal structure. Davis et al. (2018, 2020b) have recently adapted their OU network formalism to study human control. We thus see the current project as extending David et al.'s analyses of complex dynamic learning and control.

2.2 Dynamic Control

The work presented here connects closely to the literature on control behaviour in complex dynamic environments. Control has been extensively studied in a variety of tasks such as complex problem-solving tasks (Burns and Vollmeyer, 2002), computer-simulated scenarios (Brehmer and Dörner, 1993), dynamic decision-making tasks (Berry and Broadbent, 1984), microworlds (Brehmer, 2005) or naturalistic decision-making tasks (Lipshitz et al., 2001), and spanning several research domains such as economics, engineering or human–computer interaction. Here, we follow Osman (2010) and refer to the tasks as complex dynamic control (CDC) tasks.

Two factors CDC tasks have in common are that they study sequential decision-making in systems whose dynamics are hidden or initially unknown, and that these systems change autonomously or as a direct consequence of our actions (Brehmer, 1992). The tasks were primarily developed to evaluate people´s ability to successfully control and make effective decisions in stylised, and often ill-defined, simulations of complex real-world scenarios involving uncertainty, feedback loops, multiple goals or multiple agents (Funke, 2001). However, a large body of research has started to unpack the various psychological processes involved in performing control tasks such as attention, implicit learning, memory, monitoring or planning (for a review, see Osman, 2010).

Theoretical explanations of the cognitive processes underlying control behaviour put different weight on the extent to which people form and rely on abstract representations of the task structure and parameters. Berry and Broadbent (1984, 1987), for example, explored a task where people have to control the size of the workforce in a sugar factory to achieve a specific level of production. They found that task performance was not linked their ability to explicitly learn the structure underlying the task. These findings prompted the development of instance-based learning theories (Berry and Broadbent, 1987), suggesting that people store task-specific properties at each task encounter and use them to address similar situations, rather than focusing on task structure. In contrast, Hagmayer et al. (2010) examined control behaviour within a system defined by a simple causal structure. They found that when parts of the structure were changed people responded in a way that suggests they had acquired knowledge of the causal structure itself.

Most studies, including those presented so far, have investigated control behaviour in static environments. This is because isolating the influence of structure learning on control performance requires a formal specification of the properties and structure that generate the task dynamics. This task has historically been challenging, especially in more complex control environments (Kirlik, 2007). Similarly, other studies that examine more complex dynamics without formalising their properties have struggled to determine whether control performance stems from the peoples' ability to learn structure representations or the controllability of the task itself (Lipshitz and Strauss, 1997). One exception is the work of Davis et al. (2018, 2020b), who use the OU network formalism to define the dynamics of a complex dynamic control environment. In both of their papers, the authors conduct a systematic analysis of the conditions in which people develop causal structure representations, finding that people were best described by a simple control strategy that uses minimally complex structure representations.

We differ from the work of Davis et al. in two respects. First, while the OU network formalism permits the modelling of complex phenomena, their control task did not require participants to engage with the full complexity of the task environment, making minimal structure representations sufficiently effective. This paper considers a more challenging control task involving two control variables (rather than one) and discrete intervals (instead of continuous time). With these modifications, we aim to introduce a level of realism that engages the complex problem-solving behaviour we are interested in. Second, Davis et al.'s models were highly correlated. In our task, the incorporation of discrete time intervals and multiple control variables leads to a significantly expanded set of feasible actions, enabling a more nuanced investigation of goal-directed behaviour.

Chapter 3

Framework

In this paper, we present a dynamic control task in which participants have to adjust two sliders to place a target variable in a target region without knowing the process that generates the movement of the target variable. A reward function provides a scalar value as feedback, where positive rewards are given when the target variable´s value falls within the target region, encouraging strategic adjustments of the two sliders.

The transition dynamics of the task are defined by a stochastic Markov decision process (MDP), consisting of the two slider variables, A and B, and the target variable, C, which change their state in discrete time steps. The three variables are causally related and their interdependencies are determined by a causal structure. The structure between the variables can be illustrated by a causal graph representing variables as nodes and their conditional dependencies, i.e. causal relationships, as edges between nodes (see Fig. 3.1).

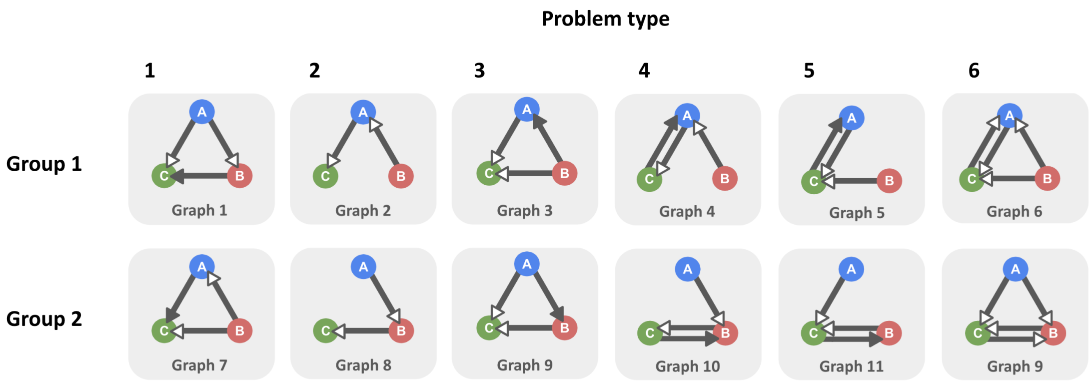

Figure 3.1: Causal graphs examined in the experiment. All problem types were presented in random order and counterbalanced between Groups 1 and 2. Regular and inverse causal connections are depicted using black and white arrowheads respectively.

To model how the dynamics of a causal structure unfold in time, we follow Davis et al. (2018, 2020b) and define the state transitions of the variables based on Ornstein–Uhlenbeck (OU) processes. OU processes are a type of stationary Gauss-Markov process developed in the field of physics to describe the stochastic movement of a variable toward a stable mean over time (Uhlenbeck and Ornstein, 1930). In the context of control tasks, this is a convenient mathematical representation of a system´s state because of the Markov property which ensures that the change of a variable depends only on its current state. This property greatly simplifies the simulation of the control environment and the modelling of different control strategies that interact with the environment. Equation 3.1 expresses the change in the state of variable i, Δv_i^t, according to the OU process.

$$P(\Delta\, v_i^t|\omega, \mu_i, v_i^t, \sigma) = \omega[\mu_i - v_i^t] + N(0,\sigma) \tag{3.1}$$

Here, ω determines how fast the variable returns to the mean, μ_i is the mean that variable i reverts to, v_i^t is the value of variable i at time t and σ defines its variance.

The standard OU process can be modified to describe the causal influence of a set of interdependent variables. Specifically, we assume that, when a variable is causally related to others, it reverts to a non-stationary mean determined by a linear function of the current states of its causes. Here, the change in the variable i, Δv_i^t, is given by,

$$P(\Delta\, v_i^t|v^t, \omega, \sigma, \theta_{i.}) = \omega\left[\left[\sum_j \theta_{ij} * v_j^t\right] - v_i^t\right] + N(0,\sigma) \tag{3.2}$$

where θ_{ij} expresses the strength and existence of a causal connection between variable i and each of the other variables j, and v_j^t is the value of variable j at time t. Formulated for each variable within a causal structure, the set of interrelated OU processes, the OU network, represents all causal relationships between the variables and, thus, fully describes the dynamics of a causal system unfolding in time.

By adjusting the causal structure of the variables OU networks are able to represent complex dynamics, including feedback loops and oscillations. Some of these more complex relationships are explored in problem types four to six in this study, which introduce feedback (see Fig. 3.1). A more detailed exploration of the emergent behaviours within OU networks is provided by Davis et al. (2020a).

Finally, note that OU networks are of course not the only way to express the dependencies and interactions between causally related variables over time (e.g., Griffiths and Tenenbaum, 2009; Pacer and Griffiths, 2011). More generally, they can be considered a specific instance of a broader probabilistic modelling framework, Dynamic

Bayesian Networks (DBNs), in which the functional forms describing the relationships between variables and defining how multiple causal influences combine are specific to OU networks (Koller and Friedman, 2009).

Chapter 4

Modelling

We now present three computational models of the cognitive processes underlying human goal-directed learning and explore their qualitative predictions of causal learning in the above control task. All models share the assumption that successful control involves forming a representation of the causal structure generating the task dynamics. However, the models differ in their predictions regarding the extent and complexity of those representations, spanning from optimal learning (Causal Model Based Controller) to heuristic mappings between variables (Local Computations Controller) and actions and outcomes (Proportional-Integral-Derivative Controller). Moreover, the models choose to apply their knowledge in different ways. While the first two use their representations to project the effect of the different action choices into the future, the third model employs its representation to steer actions based on a historical record of past errors.

4.1 Causal Model Based Controller

The Causal Model Based Controller (CMBC) (Davis et al., 2018, 2020b) gives a normative account of goal-directed learning. Specifically, it predicts that individuals learn to control the task as if they are guided by a distributional belief over all potential causal structures underlying the task environment. Whenever new evidence is observed, the CMBC updates its estimate of the posterior distribution over causal graphs. A causal graph, $g \in G$, defines for every potential causal connection between two variables x_i and x_j whether it is regular ($\theta_{x_i x_j} = 1$), inverse ($\theta_{x_i x_j} = -1$), or not existing ($\theta_{x_i x_j} = 0$). For a causal structure with three variables, A, B and C, this representation allows for six potential causal connections per graph, denoted as $\theta_g = [\theta_{AB}, \theta_{AC}, \theta_{BA}, \theta_{BC}, \theta_{CA}, \theta_{CB}]$. As a result, the hypothesis space considered for the task in this paper contains $3^6 = 729$

possible causal graphs.

To compute the posterior probability of a particular causal graph, the CMBC calculates the likelihood of the observed change in each variable and at each time step. This is done considering the values of all variables at time t, as well as the specific values of the causes of x_i that are associated with the causal graph. We define the likelihood as,

$$P(\Delta v^t_{x_i}|v^t, \omega, \sigma, \theta_{x_i \cdot, g}; I^t_i) = \begin{cases} 1 & \text{if } I^t_i \\ N(\omega(\sum_j \theta_{x_i x_j, g} v^t_{x_j} - v^t_{x_i}), \sigma) & \text{otherwise} \end{cases} \tag{4.1}$$

where $\Delta v^t_{x_i}$ is the likelihood of the observed change in variable x_i at time step t, v^t holds the observed values of all variables at time t, $\theta_{x_i \cdot, g} \subset \theta_g$ is the set of causes of x_i that are associated with graph g, and ω and σ are the system parameters. The likelihood of the observed change in a variable takes into consideration potential interventions on the variable, as denoted by a binary indicator variable I_{ti}. The indicator variable takes the value true (1) if the variable has been intervened on, and false (0) otherwise. Multiplying the likelihood of all variables and for each previous time step gives the joint likelihood of all observed variables at all time points,

$$P(\mathbf{v}|g, \mathbf{I}) = \prod_{i=1}^{N} \prod_{t' \in T} P(\Delta v^{t'}_{x_i}|v^{t'}, \omega, \sigma, \theta_{x_i \cdot, g}; I^{t'}_i) \tag{4.2}$$

Here, $\mathbf{v}$ is the vector of observed values of all variables up to t, g is the causal graph, $\mathbf{I}$ is the set of interventions up to t, T denotes the set of all previous time steps up to time t and N is the total number of variables. To obtain the posterior probability of the causal graph, the likelihood is multiplied by the prior probability of the graph (uniform distribution) and then normalised by the evidence, as expressed in Equation (4.3).

$$P(\mathbf{v}|g, \mathbf{I}) = \frac{P(\mathbf{v}|g, \mathbf{I})P(g)}{\sum_{g \in G} P(\mathbf{v}|g, \mathbf{I})P(g)} \tag{4.3}$$

The CMBC uses its best estimate of the causal structure to choose the action that maximises immediate reward. For this purpose, the model engages in planning; it selects the causal graph with the highest posterior probability and projects the effect of each possible action choice on the control outcome several time steps into the future. The expected reward of a particular action choice is determined based on integrating the distribution of the target variables' projected values that fall within the target region. Equation (4.4) defines the expected value of a particular action choice a at time t as the sum of discounted expected rewards.

$$EV_t(a|g^*) = \sum_k \gamma^k \int_{r_{min}}^{r_{max}} N(\mu_T; s) dx \tag{4.4}$$

Here, $g*$ is the graph with the highest posterior probability, k is the number of projected time steps, γ is a discount factor, r_{min} and r_{max} define the limits of the target region and μ is the simulated mean state of the target variable. The locally optimal action at time t is that which maximises the expected value,

$$A_t \doteq \arg\max_a EV_t(a|g^*) \tag{4.5}$$

For the task in this paper, we project the impact of a choice for three consecutive time steps and discount it with $\gamma = 0.9$.

4.2 Local Computations Controller

The Local Computations (LC) Controller (Davis et al., 2018, 2020b) offers a computationally less demanding account of goal-directed learning compared to the CMBC. Inspired by the local computations heuristic, the LC Controller operates under the assumption that people simplify complex causal learning problems and only consider causal connections between pairs of variables (Fernbach and Sloman, 2009). Accordingly, the model estimates the causal connections between two variables considered in isolation rather than estimating the entire causal structure at once. Similar to the CMBC, the LC Controller uses its pairwise estimates to simulate potential future trajectories for all possible actions and determine the optimal course of action. We can formalise the LC Controller by adapting Equation 4.1 as follows,

$$P(\Delta v^t_{x_i} | v^t, \omega, \sigma, \theta_{x_i \cdot, g}; I^t_i) = \begin{cases} 1 & \text{if } I^t_i \\ \sum_j N(\omega(\theta_{x_i x_j, g} v^t_{x_j} - v^t_{x_i}), \sigma) & \text{otherwise} \end{cases} \tag{4.6}$$

Although pair-wise causal structure representations are very efficient, they introduce a systematic bias when two variables are mediated by a third (Davis et al., 2020a; Rehder et al., 2022). Consider, for example, the causal chain $A \to B \to C$ (θ_{BA}, $\theta_{CB} = 1$, θ_{AB}, θ_{AC}, θ_{BC}, $\theta_{CA} = 0$). While A and C are not directly causally related, changes in A will indirectly influence the value of C, resulting in a correlation between the two variables. The CMBC would correctly infer no direct connection between A and C because it considers all causal influences of C simultaneously. However, the LC is prone to infer $A \to C$ in addition to $A \to B$ and $B \to C$ because it cannot account for the mediating role of B by considering causal links in isolation.

This behaviour produces a testable claim that can be compared to human causal judgements. Figure 4.1 (B) displays the accuracy of the models in detecting causal connections based on link types across all the problem types examined in this paper. The LC model performs worse than the CMBC at identifying indirectly mediated links (e.g., θ_{AC} in the above example), while the models are on par for direct links (e.g., θ_{AB} and θ_{BC}).

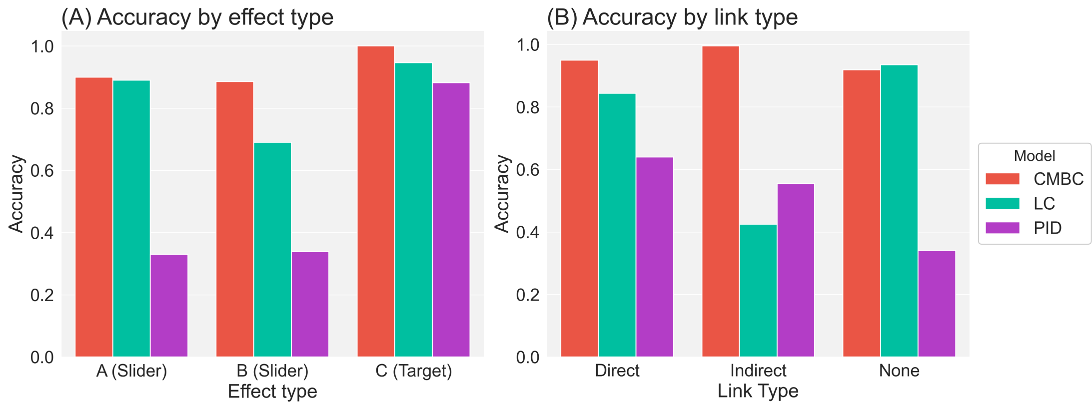

Figure 4.1: Accuracy of the computational models at identifying causal connections by effect type (A) and link type (B) assuming the most probable graph is selected at the end of each trial. Results are averaged over 100 simulations of each problem type used in the experiment (see Fig. 3.1).

4.3 Proportional-Integral-Derivative Controller

The Proportional-Integral-Derivative (PID) Controller offers the most heuristic account of goal-directed learning out of the three models. Originally developed for industrial process control, the algorithm has been found to align well with how people generate actions in dynamic control tasks (Ritz et al., 2018; Davis et al., 2018). It predicts that people only consider the effect of their actions on the control outcome and adjust them based on the discrepancy between desired and actual outcomes. Equation 4.7 captures the error (e_t) between the mean target region and the state of the target variable (v_C^t), where r_{min} and r_{max} define the limits of the target region.

$$e_t = \frac{r_{min} + r_{max}}{2} - v_C^t \tag{4.7}$$

To bring the target variable closer to the desired target region, the PID Controller selects a corrective action based on a weighted combination of the current error (P), accumulated error over time (I), and the rate of error change (D) as shown in Equation (4.8).

$$u_t = K_P e_t + K_I \sum_{n=1}^{t} e_n + K_D (e_t - e_{t-1}) \tag{4.8}$$

Here, K_P, K_I, and K_D denote the coefficients for the proportional, integral, and derivative terms of the PID Controller, respectively.

To achieve meaningful corrective actions, the PID Controller considered in this paper estimates the causal connections between sliders A and B, and the target variable C. Other causal links are assigned randomly. In case the PID Controller predicts that both sliders are causes of the target variable it randomly chooses which slider to intervene on. Equation 4.9 defines the corrective action given the weighted error (u_t), and previous state ($v_{x_j}^{t-1}$).

$$A_t \doteq \theta_{Cx_j} u_t + v_{x_j}^{t-1} \tag{4.9}$$

Here, θ_{Cx_j} denotes the weight of the slider that the PID Controller intervenes on. Importantly, this behaviour predicts that people deviate systematically from normative causal inferences. Figure 4.1 (A) illustrates the accuracy of the different models in identifying causal connections by effect type, based on 100 simulations spanning all problem types. In contrast to the CMBC and LC Controller, the PID Controller identifies only the outgoing causal links from slider variables A and C to the target variable C at above chance level.

Chapter 5

Method

5.1 Participants

100 participants (38 female, age M = 40.42, SD = 11.27) were recruited from Prolific, a crowdsourcing platform for research. They were paid a base payment of £2.5 and received additional performance-related bonus payments (M = £0.56, SD = £0.21). The experiment took on average 20 minutes (SD = 9.40). In a post-task questionnaire, participants rated the control task as engaging (M = 8.24, SD = 1.79) and difficult (M = 8.68, SD = 1.61) on a ten-point scale ranging from 1 (lowest) to 10 (highest). All procedures were approved by the School of Philosophy, Psychology & Language Sciences Research Ethics Committee (PPLS REC).

5.2 Materials

Participants interacted with three variables, A, B, and C, represented by line graphs on a dynamic chart (see Fig. 5.1). The variables changed their value in discrete time steps following the OU process and adhering to a causal structure that determined their interdependencies. Two of the variables, A and B, were additionally presented as vertical sliders. The sliders could be interacted with, by clicking, dragging and holding their handle to adjust the values of variables A and B and override the state of the OU process. The slider values were rounded to integers and restricted within the range of -100 and 100. The goal of the task was to place the third variable, C, in a target region between the values of 35 and 65 that was coloured in yellow. A timer at the top of the page recorded the steps taken, from 0 to 40, at an increment of two seconds.

On completion of a trial, six queries were presented, each corresponding to a potential causal relationship between two of the three variables (see Fig. 5.2). The responses were classified as 'Regular', 'Inverse' or 'None', equivalent to an attracting relationship ($\theta > 0$), repelling relationship ($\theta < 0$), or no causal relationship ($\theta = 0$) between two variables respectively. Each query was accompanied by a confidence rating on a seven-point scale ranging from 1 ('not confident') to 7 ('very confident').

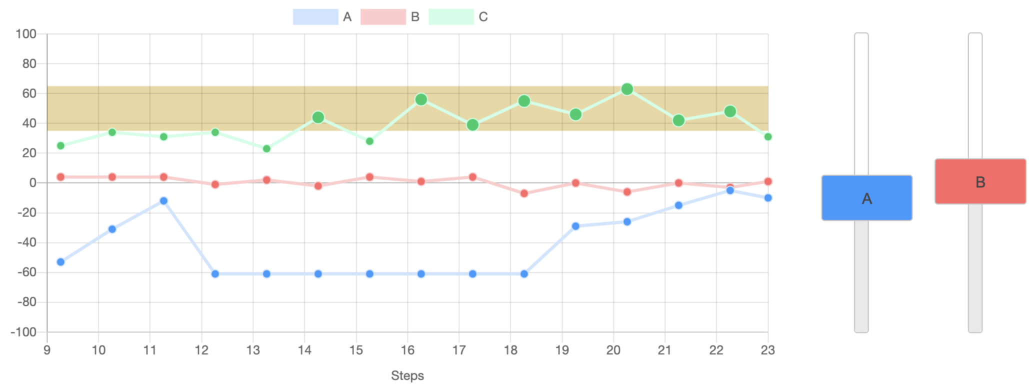

Figure 5.1: The dynamic control task as it is presented on a computer screen. The dynamic chart with three line graphs is updated in discrete intervals and according to the OU process and causal graphs. Control variables, A and B are presented as sliders. The slider values, ranging from -100 to 100, are by default updated as a function of the states of their causes or upon participant intervention.

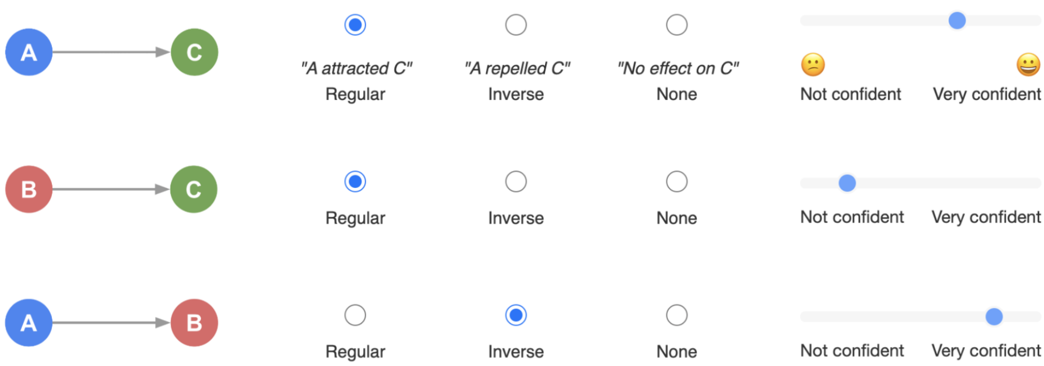

Figure 5.2: Excerpt of the causal queries following each control task. Not shown here are the queries on $B \rightarrow A$, $C \rightarrow A$ and $C \rightarrow B$.

5.3 Stimuli and Design

Participants interacted with six causal graphs presented in random order over the course of six trials. The causal graphs consisted of common causal networks, including chain and common cause networks, as well as, more complex networks including feedback loops (see Fig. 3.1). All causal networks were chosen such that they exhibit a direct 'Inverse' link, $\theta = -1$, from one of the controlled variables, A and B, to the target variable C. The structure of the network was systematically varied across trials to mitigate any potential effects or preferences associated with the position of the sliders or the order of the slider manipulation (cf. Group 1 and 2, Fig. 3.1). The parameters of the OU process were set to $\omega = 0.5$ and $\sigma = 5$. The causal relationships were limited to $\theta = [-1, 0, 1]$ for a repelling relationship, no relationship, or attracting relationship respectively. Slider values were updated every two seconds in accordance with the OU process.

5.4 Procedure

After giving their informed consent (see Appendix C), the participants completed an instruction phase informing them about the nature, duration and control goal of the task. The instruction phase included a series of slides and a 40-second video demonstration of the task showing each slider being manipulated twice, by clicking and dragging the slider handle. Participants needed to successfully complete a comprehension quiz consisting of four questions to proceed to the task. The purpose of this quiz was to make sure participants had understood (1) how to manipulate variables A and B through slider intervention, (2) that there exist conditional dependencies between the variables, (3) the relationship between placing variable C in the target region and earning bonus payment, and (4) the goal of the task of maximising their bonus pay. They had the opportunity to review the instructions and retake the quiz until all questions were answered correctly.

In the main phase of the experiment, participants completed six trials of the control task, with each trial lasting for 40 steps (i.e. 80 seconds). To initiate a trial, participants pressed a "Start" button located at the bottom of the page, triggering the variables to start updating according to the OU process and in 2-second intervals. Participants were able to manipulate any slider by clicking, dragging, or holding it at any desired value, thereby overriding its behaviour within the OU process for that specific step. Once intervened upon, the other slider was temporarily disabled for that step. Upon releasing a slider, the variable resumed its movement in accordance with the OU process on the next step. Each step that variable C was successfully placed in the target region the bonus pay increased by approximately £0.01, and until it reached a maximum bonus pay of £1.50. At the end of each trial, participants were presented with six causal queries and were required to enter a judgment and confidence rating for all six potential causal relations to proceed to the next trial. Participants were not informed about the causal queries in the instructions ahead of the task.

Following the completion of all six trials, participants completed a short post-task questionnaire, which included a free-text prompt to describe the strategies they used to solve the tasks.

Chapter 6

Results

6.1 Controllability

Across all six problem types, participants scored on average 7.83 (SD = 5.48) points out of a possible 40 points in the task, which is significantly higher than what would be expected by chance alone, t(99) = 9.69, $p < 0.01$. They also significantly improved control performance within a trial, as shown by a linear regression of the proportion of participants receiving rewards on the number of steps taken, b = 0.005, t(244) = 20.06, $p < 0.01$ (see Fig. 6.1 (A)). Participants were more likely to score rewards as the experiment progressed, as demonstrated by a linear mixed model analysis of the proportionate rewards earned on trial number and with subject-level intercept (b = 0.37, t(99) = 2.96, $p < .01$, $R^2 = 0.62$). The results suggest participants were able to transfer task-related knowledge to subsequent trials.

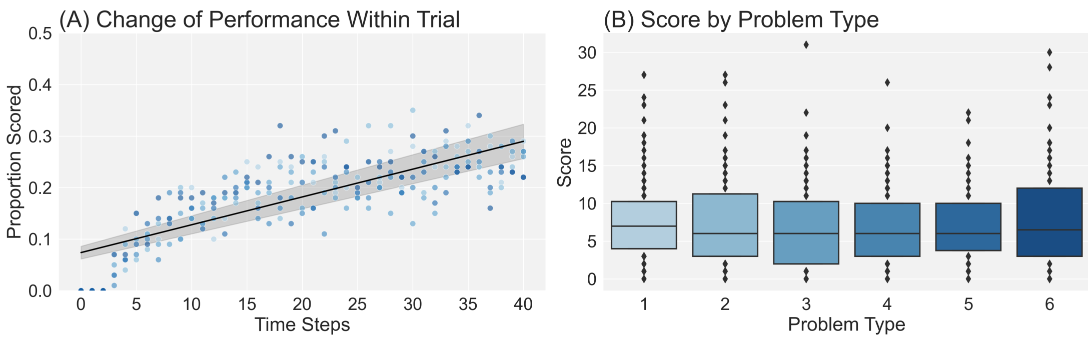

Figure 6.1: (A) Proportion of participants receiving a reward (scoring) per time step, with each data point colour-coded based on the problem type (cf. (B) for reference). (B) Boxplot of participant scores by problem type.

There was no significant difference in mean performance between the causal graphs in Group 1 (M = 7.45, SD = 5.71) and Group 2 (M = 7.45, SD = 5.49), $t(595) = 0.004$, $p = 0.99$, indicating that there were no apparent effects associated with the positioning of the two sliders (cf. Group 1 and 2, Fig. 3.1). Therefore, we present the data collapsed over groups in the remaining analysis. Moreover, a one-way repeated measures ANOVA was conducted to assess the effect of the problem type on the amount of reward collected (see Fig. 6.2), revealing no significant effect, $F(5, 495) = 1.59$, $p = 0.16$, generalised $\eta^2 = 0.01$. This outcome was expected, given that all graphs were intentionally designed to demonstrate a direct inverse connection from one slider to the target, resulting in similar levels of difficulty in controllability.

6.2 Learnability

On the causal queries that followed the control task, participants were slightly above chance level (0.33) at successfully identifying causal links (M = 0.39, SD = 0.10), $t(99) = 5.86$, $p < 0.01$. However, accuracy did not increase as trials progressed. This was demonstrated by a linear mixed model analysis of trial number on participants' mean causal judgment accuracy with subject-level intercepts which yielded a non-significant result, $t(99) = -0.62$, $p = 0.54$. Participants also did not identify overall causal networks (M = 0.003, SD = 0.02), i.e., all six possible causal connections in a causal graph, significantly above chance ($3^6 = 0.0014$), $t(99) = 0.84$, $p = 0.20$. The participants' relatively poor performance was reflected in their confidence ratings. Overall, participants reported low confidence in their judgments (M = 2.82, SD = 1.87) on a confidence rating from 1 ('not confident') to 7 ('very confident').

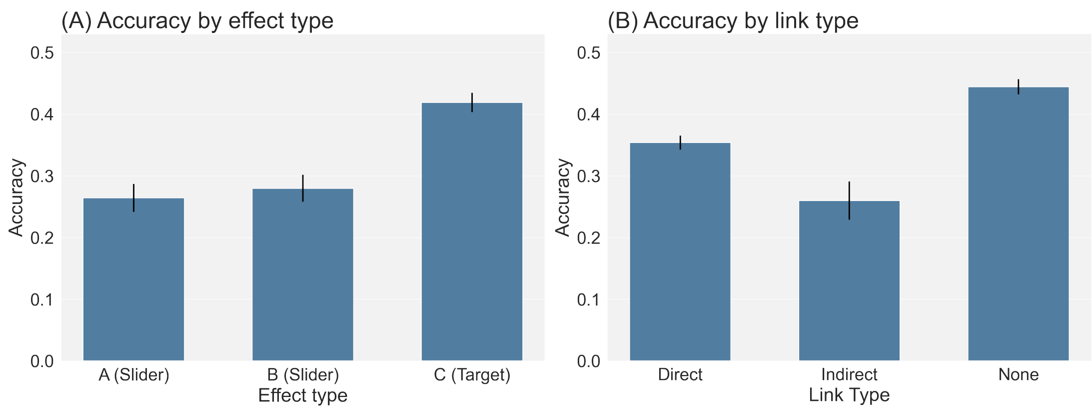

Figure 6.2: Participants' accuracy in identifying causal links by (A) causal effect type and (B) causal link type. Error bars represent standard errors of the mean.

Consistent with the theoretical predictions of the PID Controller presented earlier, participants were slightly more likely to identify existing causal connections in cases where the sliders influenced the target variable (M = 0.42, SD = 0.18), compared to cases that affected the slider variables (M = 0.27, SD = 0.19), $t(99) = 7.06$, $p < 0.01$ (see Fig. 6.2(A)). However, their responses did not reveal a preference for simpler models. Participants considered over half of the possible links to be causally related (M = 0.62, SD = 0.23), a proportion significantly greater than the true ratio of causal links in the underlying networks (0.50), $t(99) = 5.00$, $p < 0.01$.

Accuracy also varied with the type of causal connection (see Fig. 6.2(B)). Participants were more likely to correctly identify direct (M = 0.35, SD = 0.15) or absent links (M = 0.44, SD = 0.27) compared to indirect links (M = 0.26, SD = 0.33), i.e., those links where two unrelated variables are correlated due to an interaction with a third variable, $F(2,198) = 11.64$, $p < 0.01$. The results align with the predictions of the LC Controller, suggesting that people often erroneously infer a direct causal relationship when two variables are only indirectly related (e.g., A and C with $\theta_{CA} = 0$ in $A \rightarrow B \rightarrow C$). It is important to note, however, that the strength of this signal is relatively weak compared to previous findings in causal learning tasks (Rehder et al., 2022).

6.3 Interventions

To succeed in the control task, effective interventions are essential. Here, we define an intervention as the participant's active choice to interact with the task environment during each time step, instead of passively observing the task dynamics unfold. We recorded interventions whenever the participant adjusted either of the two sliders by clicking, holding, or dragging their handles, resulting in 18,838 interventions across trials.

Throughout the trials, participants allocated a significant portion of their time to interventions, spending on average 78% (SD = 16%) of the trial time intervening on the sliders. The distribution was slightly skewed, with a median of 83% and a mode of 93%, mainly due to a few participants who only made a small number of interventions across trials. The average time spent on interventions per trial slightly increased as the experiment progressed, $b = 0.44$, $t(99) = 4.00$, $p < 0.01$.

Participants used both sliders extensively during a trial, distributing their interventions evenly between sliders A (M = 40%, SD = 14%) and B (M = 41%, SD = 14%), $t(99) = -1.04$, $p = 0.30$. The majority of interventions involved adjusting a slider for a single step followed by either inaction (M = 22%, SD = 14%) or intervening on the respective other slider (M = 42%, SD = 18%). In cases where participants intervened on the same slider for more than a single time step (M = 36%, SD = 16%), their interventions lasted on average three time steps (SD = 1.78) and displayed considerable variability in the range between the sliders' maximum and minimum values (M = 64.84, SD = 54.00). The observed 'juggling' behaviour, wherein participants alternate between sliders, was unexpected considering that the causal graphs for all problem types (except for problem type 6, as illustrated in Fig. 3.1) were designed to allow the control goal to be reached through intervention on either slider individually.

We observed a degree of systematicity in participants' slider adjustments during an intervention. The average values of the sliders after interventions significantly differed from what would be expected by purely random behaviour (with the exception of slider variable B in problem type 2). Additionally, participants' slider adjustments showed a slight tendency toward optimal behaviour. For a detailed illustration of how participants' average slider adjustments compare to optimal behaviour, please refer to Appendix A, Figure A.1. Moreover, interventions were moderately correlated. A linear regression analysis revealed a significant positive relationship (b = 0.60) between the value of an intervention and that of the preceding intervention for each slider ($t(16{,}955) = 92.85$, $p < 0.01$), indicating that participants adjusted their interventions based on the outcomes of previous interventions (see Appendix A, Figure A.2).

6.4 Results Summary

Overall, participants recognised the significance of interventions for task control. They understood the utility of adjusting both sliders in service of the control goal and the varying effectiveness of different slider positions. However, their ability to intervene effectively and learn the task's causal structure was only moderate. While participants displayed systematic causal judgment errors as anticipated by the LC and PID Controllers, the signal strength of this behaviour was relatively weak. Thus, by inspection of the data alone, it remains uncertain to what extent participants relied on abstract representations of the task environment in their attempt to control the task.

Chapter 7

Model Comparison

We now evaluate the models on their ability to predict participant interventions in the experiment. Fig. 7.1 shows the average reward curves of the models when they are optimised for task performance. First, note that while the CMBC and LC Controller perform nearly optimally across all problem types, the performance of the PID Controller is noticeably lower. This decline can likely be attributed to the random component in the behaviour of the PID Controller. The strength of PID Controllers lies in their ability to provide coordinated adjustments of a single control variable. However, in this task, the model selects the slider to intervene randomly whenever it has causal associations for both sliders and the target. Introducing randomness disrupts this coordination, leading to unnecessary or conflicting adjustments that do not effectively reduce the error or achieve stability. Further, note that all three models outperform human participants, likely because they do not account for decision noise or the possibility that no action is executed. So, we proceed with a quantitative comparison of the models when they are fitted to maximise the likelihood of participant data.

The three models generate action decisions for each time step in a trial, taking as input the states of the three variables, A, B and C, and participant interventions. The CMBC and LC Controller select an action based on a vector of expected values over all action choices. To accommodate for potential decision noise when choosing an appropriate action, the vectors of expected values are passed through a softmax function, converting it into a probability distribution over action choices,

$$P(\vec{z})_i = \frac{\frac{e^{z_i}}{\tau}}{\sum_j \frac{e^{z_j}}{\tau}} \tag{7.1}$$

where $\vec{z}$ denotes the vector of expected values, z_i denotes the ith element of $\vec{z}$ and τ is

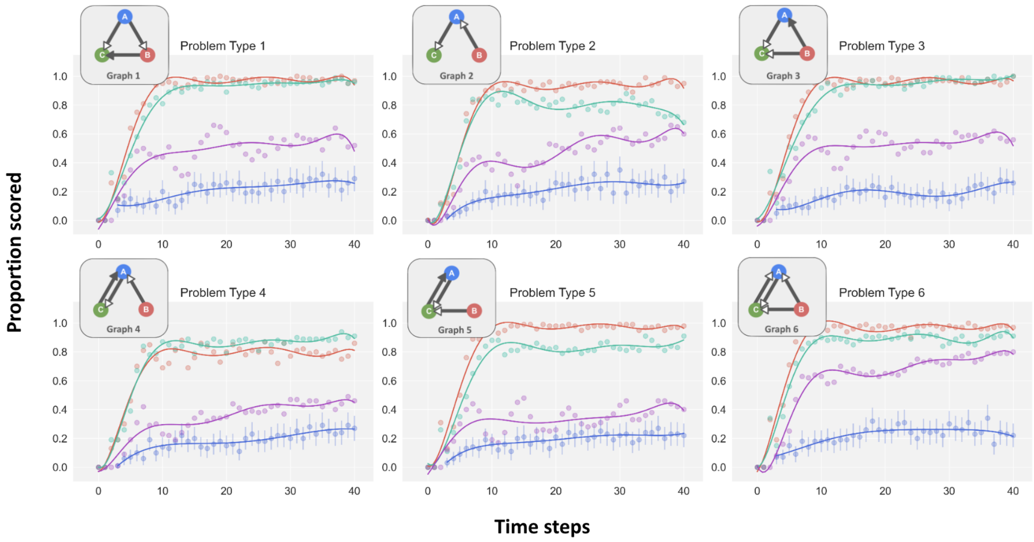

Figure 7.1: The proportion of possible reward received over the course of a trial by condition. Blue lines represent participants; red lines the CMBC Controller; green lines the LC Controller and purple lines the PID Controller. Error bars denote 95% CI (normal approximation to binomial). The data is collapsed over groups.

a temperature parameter. The temperature is fitted to maximise the log-likelihood of participant actions, adjusting the smoothness of the probability distribution. Higher values of temperature result in increased decision noise.

The PID Controller provides a single action which adjusts one of the sliders in proportion to its error. We represent the selected action (z^*) as a one-hot vector over possible values of the selected slider ($\vec{z*}$). The probability of selecting action z_i can be constructed based on two cases. In the first scenario, where only one slider is predicted to be causally relevant, the probability distribution is given by,

$$P(\vec{z})_i = \begin{cases} \dfrac{\varepsilon}{|\vec{z}|} + (1-\varepsilon)P(z_i|z^*, \sigma) & \text{if } z_i \in \vec{z*} \\ \dfrac{\varepsilon}{|\vec{z}|} & \text{otherwise} \end{cases} \tag{7.2}$$

where ε is a small value between 0 and 1, and $|\vec{z}|$ represents the total number of actions. If z_i belongs to the set of actions associated with the chosen slider $\vec{z*}$, its probability consists of a base probability, $\frac{\varepsilon}{|\vec{z}|}$, and a weighted probability derived from a Gaussian probability density function centred around the position of the one-hot vector with value 1. In the second scenario, where the model faces a choice between sliders A and B, the distribution is adjusted as follows,

$$P(\vec{z})_i = \frac{\varepsilon}{|\vec{z}|} + (1-\varepsilon)\frac{1}{2}P(z_i|z^*, \sigma) \tag{7.3}$$

Here, action z_i is chosen from the sliders, A or B, with equal probability, weighted by a Gaussian probability density function centred around each slider's chosen value. The base rate ensures every action has a minimal probability of being selected. Similarly, to the temperature parameter, the standard deviation (σ) is estimated from the data to control for decision noise.

For all models, we fit a parameter (α) that scales the probability distribution over action choices with a vector that accounts for steps where no action is executed,

$$[P(\vec{z})(1-\alpha), \alpha] \tag{7.4}$$

The three models are evaluated against a baseline which assumes each action is randomly selected from the space of possible action choices. No parameters were fitted for the baseline.

Models	NLL	BIC	τ	α	σ	ε	Mean BIC	Number Best Fit
Baseline	143,974	287,949					2879.49	1
CBMC	125,395	250,810	5018.68	0.22			2503.81	56
LC	125,435	250,890	8404.72	0.22			2505.28	39
PID	127,592	255,214		0.22	150.22	0.99	2704.39	4

Table 7.1: Results of the model comparison. All models were fitted to maximise the log-likelihood of participant actions.

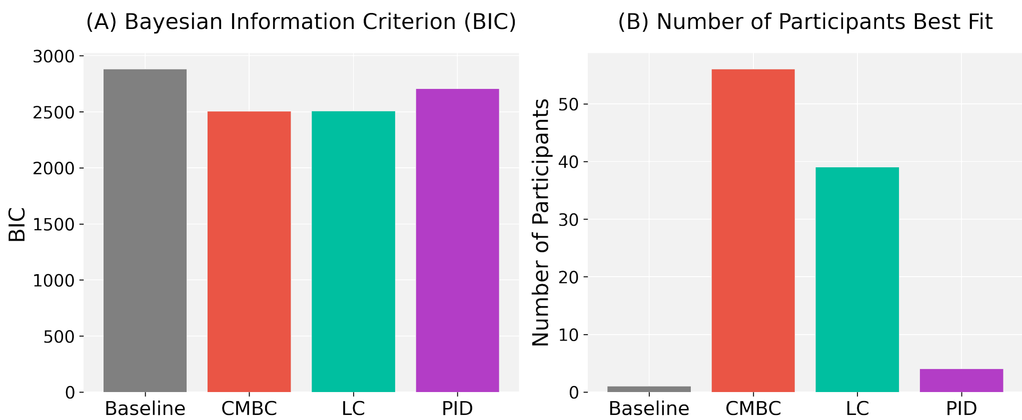

Figure 7.2: Evaluation measures for all models against the baseline. (A) Mean Bayesian Information Criterion (BIC) value per participant. (B) Number of participants best fit by each model as a function of BIC.

Table 7.1 summarises the results of the model comparison. For each model, we report the negative log-likelihood (NLL) and the Bayesian Information Criterion (BIC) at the experiment level with their associated fitted parameter values, as well as, the mean participant level BIC value (Mean BIC) and the number of participants best fit by each model (Number Best Fit). A posthoc Chi-squared test confirmed a significant deviation from a uniform distribution in the number of participants best fit across models ($\chi^2_3 = 53.03, p < 0.01$). Across all participants, the CMBC Controller had the lowest BIC score (see Fig. 7.2). On a per-participant level, 56 out of the 100 participants were best fit by the CMBC and the LC Controller gave the best account for 39 participants. Four participants were best represented by the PID Controller and one by the baseline. The results suggest that participant interventions in the task are best described by a control strategy with more complex causal structure representations, rather than relying on more heuristic representations.

Chapter 8

Discussion

In this paper, we examined human goal-directed behaviour in the context of dynamic control tasks. Our focus was on understanding the extent to which effective control relies on developing representations of the structure and dynamics of a task environment, especially, when the environment exhibits multiple complexities of real-world dynamic systems. To address this, we designed a challenging control task featuring multidimensional sequential decision-making within discrete time intervals. We found that participants displayed some of the behavioural markers linked to simpler representations of the task structure. However, when considering interventions overall, models with a more sophisticated representation provided a better description of their behaviour.

The task dynamics were generated by a new class of dynamic systems with attractive properties for the study of causal learning. Previous research either lacked precise formalisations of the task dynamics (as seen in Lipshitz and Strauss, 1997) or greatly simplified the dynamics by considering static environments (e.g., Hagmayer et al., 2010). As a result, these studies were either unable to analyse the extent to which control relies on causal representations or failed to capture the dynamics of many real-world situations adequately. Here, we were able to achieve both objectives. Firstly, by formalising the dynamics of causally related variables over time, the OU network is able to replicate some real-world complexities such as oscillations and feedback loops. Additionally, the OU network enabled us to specify different models that predict the data-generating dynamics of our task at different levels of complexity. This, in turn, allowed us to isolate causal structure learning from other aspects that contribute to control performance and conduct a formal analysis of causal structure learning in the task.

While OU networks have the potential to generate complex dynamic phenomena in theory, they must be embedded in a suitable control task to engage the type of complex problem-solving behaviour we are interested in. In previous studies, participants interacted with OU networks in continuous time through manipulation of a single slider (Davis et al., 2018, 2020b). However, this task set-up did not require them to engage with the full complexity of the task dynamics. As a result, simple strategies with minimal structure representations proved sufficient. In this project, we aimed to address these limitations by devising a more challenging control task. We introduced intervals, which provide a higher level of realism compared to a discrete task set-up while imposing a lower temporal resolution than continuous-time settings which comes at the cost of reduced information. Additionally, we added a second control variable which created a multidimensional decision-making context, prompting participants to assess the combined impact of both sliders on the target variable. We found that these changes drastically reduced the controllability and learnability of the task. At the same time, participants displayed a tendency to make systematic errors in their causal judgments. The observed errors were consistent with the predictions of more heuristic causal structure representations, suggesting they had formed representations between pairs of variables (Fernbach and Sloman, 2009; Davis et al., 2020a) and with a focus on representing actions and their outcomes (Davis et al., 2018, 2020b).

To attain a deeper understanding of causal structure learning in our task, we examined participant interventions. By introducing discrete time intervals and multiple control variables we effectively broadened the scope of possible interactions with the task compared to previous instantiations of the OU network. This notably larger set of feasible actions enables a nuanced investigation of goal-directed behaviour. For example, we found that participants alternated between the two sliders at almost every time step, even though most of our problem types allowed for simpler control strategies such as intervening on a single slider. This behaviour might stem from a belief that manipulating both sliders simultaneously provides a sense of greater perceived control or achieves a more desired outcome, even when simpler strategies might suffice. Alternatively, manipulating both sliders might also have been perceived as the only way to address the complexity of the task. However, the greater scope of actions becomes especially relevant when comparing the different models of goal-directed learning on their ability to fit participant interventions. This is because variations in the models' control strategies are more likely to stand out (Palminteri et al., 2017). Here, we found that participant interventions were best described by a normative account of control with a sophisticated causal structure representation, despite the fact that their explicit judgments had displayed some of the behavioural markers linked to simpler representations.

This project was exploratory in nature and there are several limitations that could be addressed with future experiments. Firstly, managing two sliders simultaneously likely increased cognitive load as the participants needed to process and coordinate information from both sliders. Moreover, the observed multitasking, where participants alternated between sliders, likely led to divided attention and potentially compromised the effectiveness of the participants' control strategies. Future studies could simplify the task set-up by separately exploring the effects of two control variables or discrete time intervals. Next, while the computational models presented here varied in the extent to which they rely on causal structure representations, we have not been able to successfully disintegrate their behavioural signatures in the task. Further investigations could examine heuristic strategies or processing limitations with clearly dissociable behaviours. Finally, the success of models with sophisticated representations of the task structure might be partly linked to the limitations of the PID Controller in handling tasks involving multiple control variables. For subsequent analysis, researchers may prioritise maintaining a consistent action selection mechanism to separate the impact of different task representations from the effectiveness of various action selection mechanisms.

Goal-directed behaviour and complex problem-solving are challenging to study because they rely on intricate cognitive processes that likely involve a range of cognitive capacities at once. The OU network allowed us to adapt the formal analyses of the causal learning literature to the context of control tasks and isolate the role of causal structure learning from other aspects of control performance. We believe that this formalism, alongside well-designed control tasks, provides a fruitful experimental framework to unpack the role of causal learning in dynamic control and has the potential to contribute to many disciplines that study or rely on successful control strategies such as machine learning and artificial intelligence, engineering or economics.

Bibliography

Berry, D. C. and Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, 36(2):209–231.

Berry, D. C. and Broadbent, D. E. (1987). The combination of explicit and implicit learning processes in task control. *Psychological Research*, 49(1):7–15.

Bramley, N., Gerstenberg, T., and Lagnado, D. (2014). The order of things: Inferring causal structure from temporal patterns. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 36.

Bramley, N. R., Dayan, P., Griffiths, T. L., and Lagnado, D. A. (2017). Formalizing neurath's ship: Approximate algorithms for online causal learning. *Psychological Review*, 124(3):301.

Bramley, N. R., Gerstenberg, T., Mayrhofer, R., and Lagnado, D. A. (2018). Time in causal structure learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(12):1880.

Bramley, N. R., Lagnado, D. A., and Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3):708.

Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, 81(3):211–241.

Brehmer, B. (2005). Micro-worlds and the circular relation between people and their environment. *Theoretical Issues in Ergonomics Science*, 6(1):73–93.

Brehmer, B. and Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, 9(2-3):171–184.

Burns, B. D. and Vollmeyer, R. (2002). Goal specificity effects on hypothesis testing in problem solving. *The Quarterly Journal of Experimental Psychology: Section A*, 55(1):241–261.

Coenen, A., Rehder, B., and Gureckis, T. M. (2015). Strategies to intervene on causal systems are adaptively selected. *Cognitive Psychology*, 79:102–133.

Davis, Z., Bramley, N., Rehder, B., and Gureckis, T. M. (2018). A causal model approach to dynamic control. In *CogSci*.

Davis, Z. J., Bramley, N. R., and Rehder, B. (2020a). Causal structure learning in continuous systems. *Frontiers in Psychology*, 11:244.

Davis, Z. J., Rehder, B., Gureckis, T., and Bramley, N. R. (2020b). Human dynamic control under changing goals.

Fernbach, P. M. and Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3):678.

Funke, J. (2001). Dynamic systems as tools for analyzing human judgment. *Thinking & Reasoning*, 7(1):69–89.

Griffiths, T. L. and Tenenbaum, J. B. (2009). Theory-based causal induction. *Psychological Review*, 116(4):661.

Hagmayer, Y., Meder, B., Osman, M., Mangold, S., and Lagnado, D. (2010). Spontaneous causal learning while controlling a dynamic system.

Hitchcock, C. (2018). Probabilistic causation. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.

Holland, J. H. (2014). *Complexity: A very short introduction*. OUP Oxford.

Holyoak, K. J. and Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, 62:135–163.

Kirlik, A. (2007). Conceptual and technical issues in extending computational cognitive modeling to aviation. In *Human-Computer Interaction. Interaction Design and Usability: 12th International Conference, HCI International 2007, Beijing, China, July 22-27, 2007, Proceedings, Part I*, pages 872–881. Springer Berlin Heidelberg.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT Press.

Lagnado, D. A. and Sloman, S. (2004). The advantage of timely intervention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(4):856.

Lipshitz, R., Klein, G., Orasanu, J., and Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, 14(5):331–352.

Lipshitz, R. and Strauss, O. (1997). Coping with uncertainty: A naturalistic decision-making analysis. *Organizational Behavior and Human Decision Processes*, 69(2):149–163.

Osman, M. (2010). Controlling uncertainty: A review of human behavior in complex dynamic environments. *Psychological Bulletin*, 136(1):65.

Pacer, M. D. and Griffiths, T. L. (2011). A rational model of causal induction with continuous causes. In *Advances in Neural Information Processing Systems*, volume 24, pages 1–9.

Palminteri, S., Wyart, V., and Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences*, 21(6):425–433.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Rehder, B., Davis, Z. J., and Bramley, N. (2022). The paradox of time in dynamic causal systems. *Entropy*, 24(7):863.

Ritz, H., Nassar, M. R., Frank, M. J., and Shenhav, A. (2018). A control theoretic model of adaptive learning in dynamic environments. *Journal of Cognitive Neuroscience*, 30(10):1405–1421.

Rothe, A., Deverett, B., Mayrhofer, R., and Kemp, C. (2018). Successful structure learning from observational data. *Cognition*, 179:266–297.

Simon, H. A. (1962). *The architecture of complexity*.

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives*. Oxford University Press.

Sloman, S. A. and Lagnado, D. (2015). Causality in thought. *Annual Review of Psychology*, 66:223–247.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022):1279–1285.

Uhlenbeck, G. E. and Ornstein, L. S. (1930). On the theory of the brownian motion. *Physical Review*, 36(5):823.

Waldrop, M. M. (1993). *Complexity: The emerging science at the edge of order and chaos*. Simon and Schuster.

Appendix A

Participant Interventions

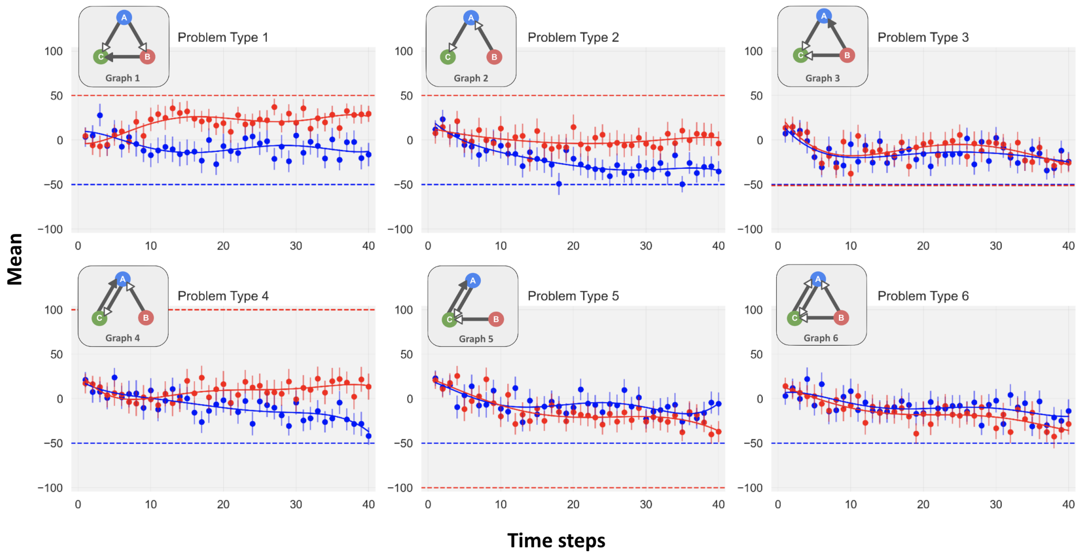

Figure A.1: Participants' average slider adjustments over the course of a trial, categorised by problem type. Blue lines represent adjustments made on Slider A, red lines represent adjustments on Slider B. Optimal adjustments based on a complete understanding of the problem type's causal structure are illustrated by dashed lines. Error bars denote 95% CI (normal approximation to binomial). The data is collapsed over groups.

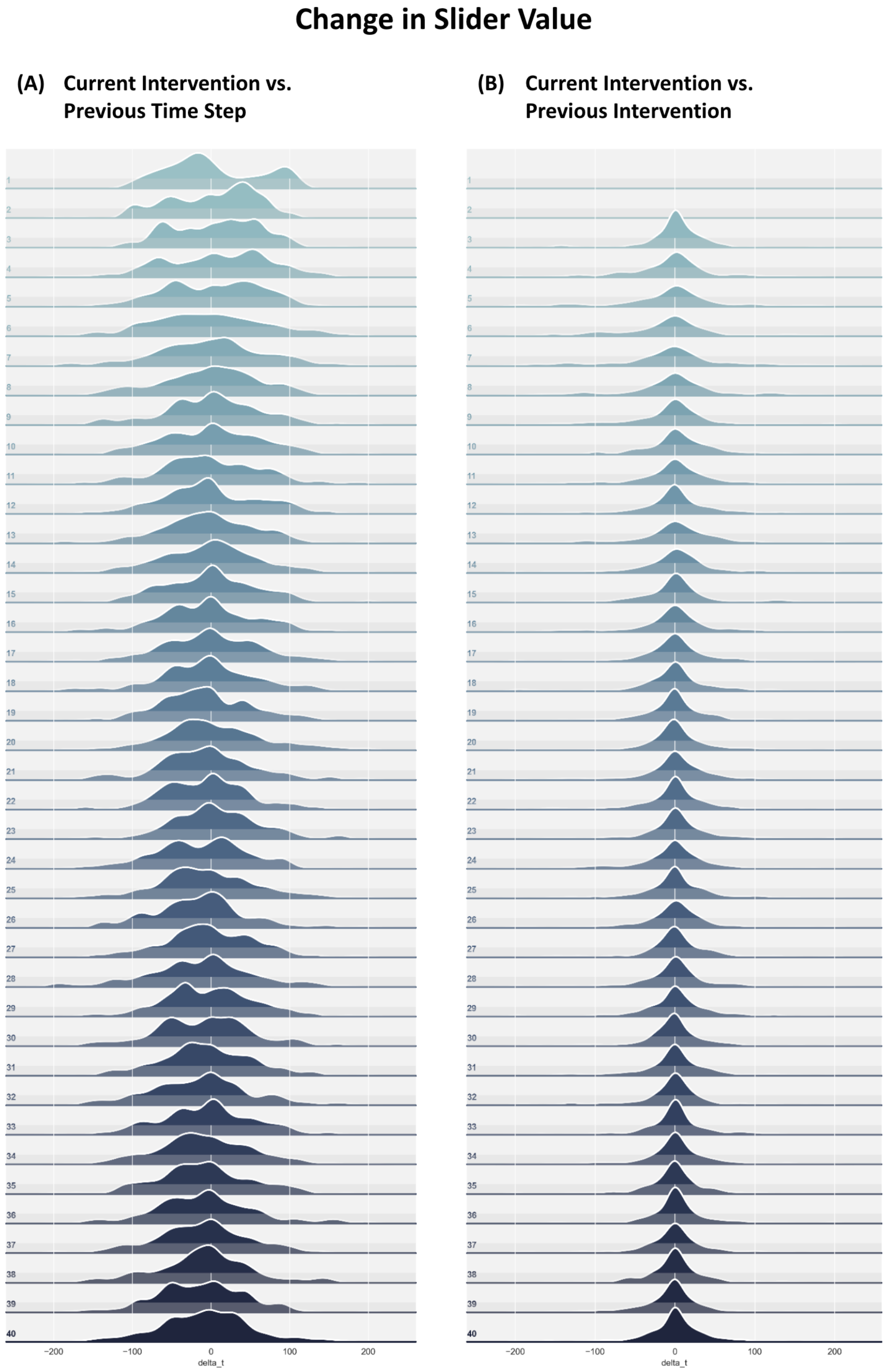

Figure A.2: Distribution of the change in slider value after intervention by time step. The change is computed with respect to (A) the slider value on the previous time step and (B) the slider value after the previous intervention on the slider. Results are presented collapsed over both sliders.

Appendix B

Participant Response Profiles

Figure B.1: Participant Response Profiles (Condition = 1 and Group = A). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

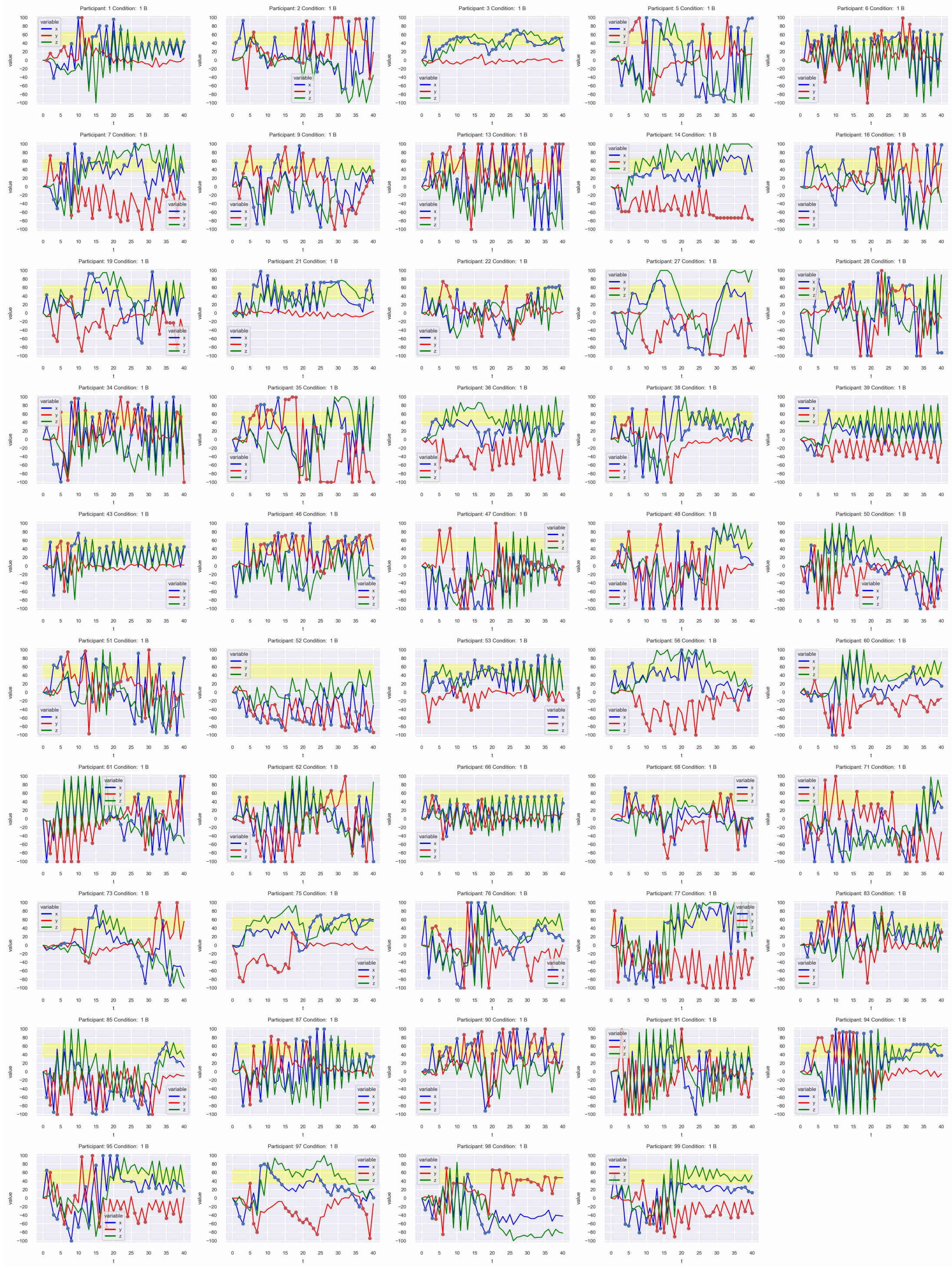

Figure B.2: Participant Response Profiles (Condition = 1 and Group = B). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

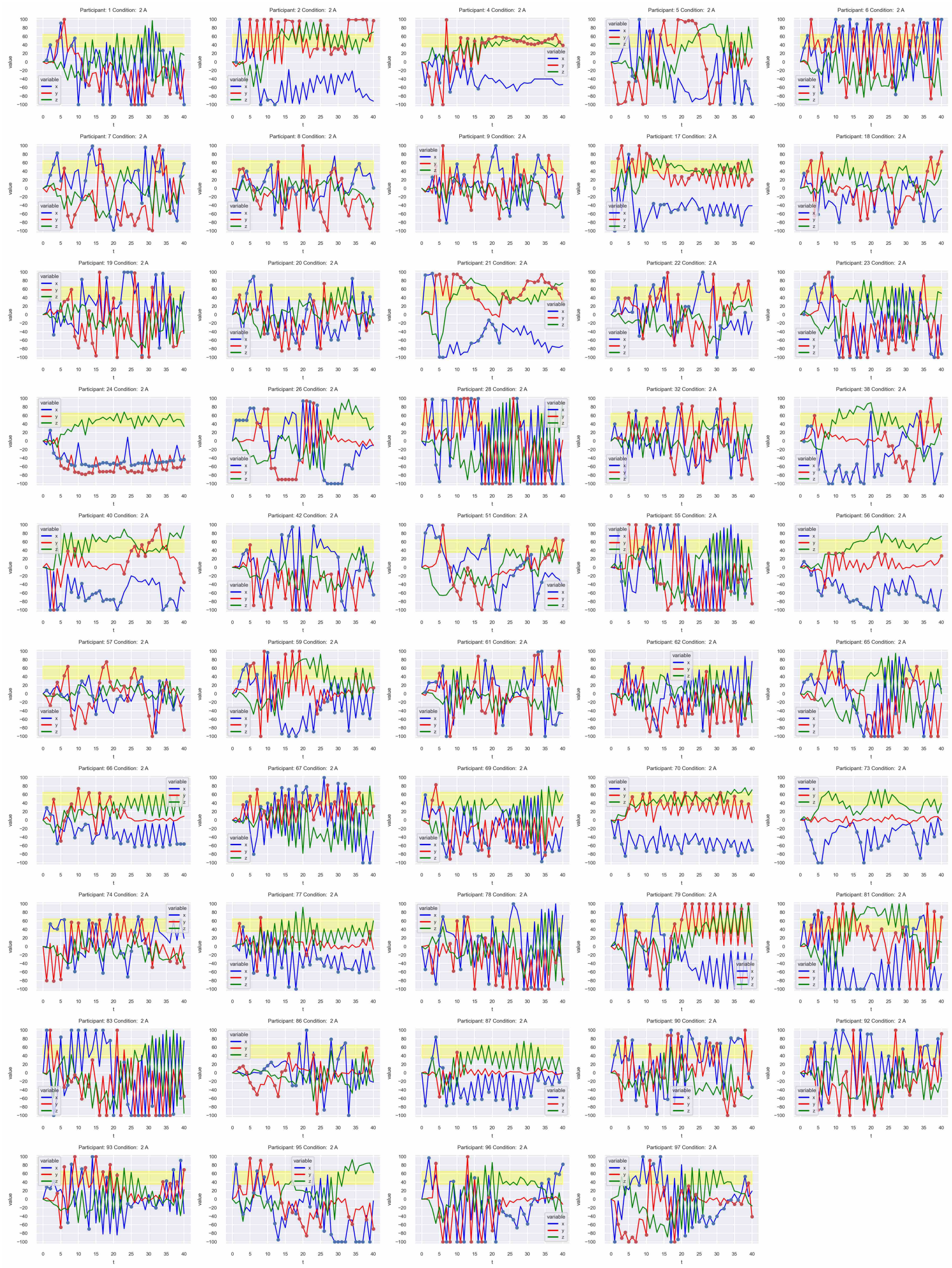

Figure B.3: Participant Response Profiles (Condition = 2 and Group = A). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

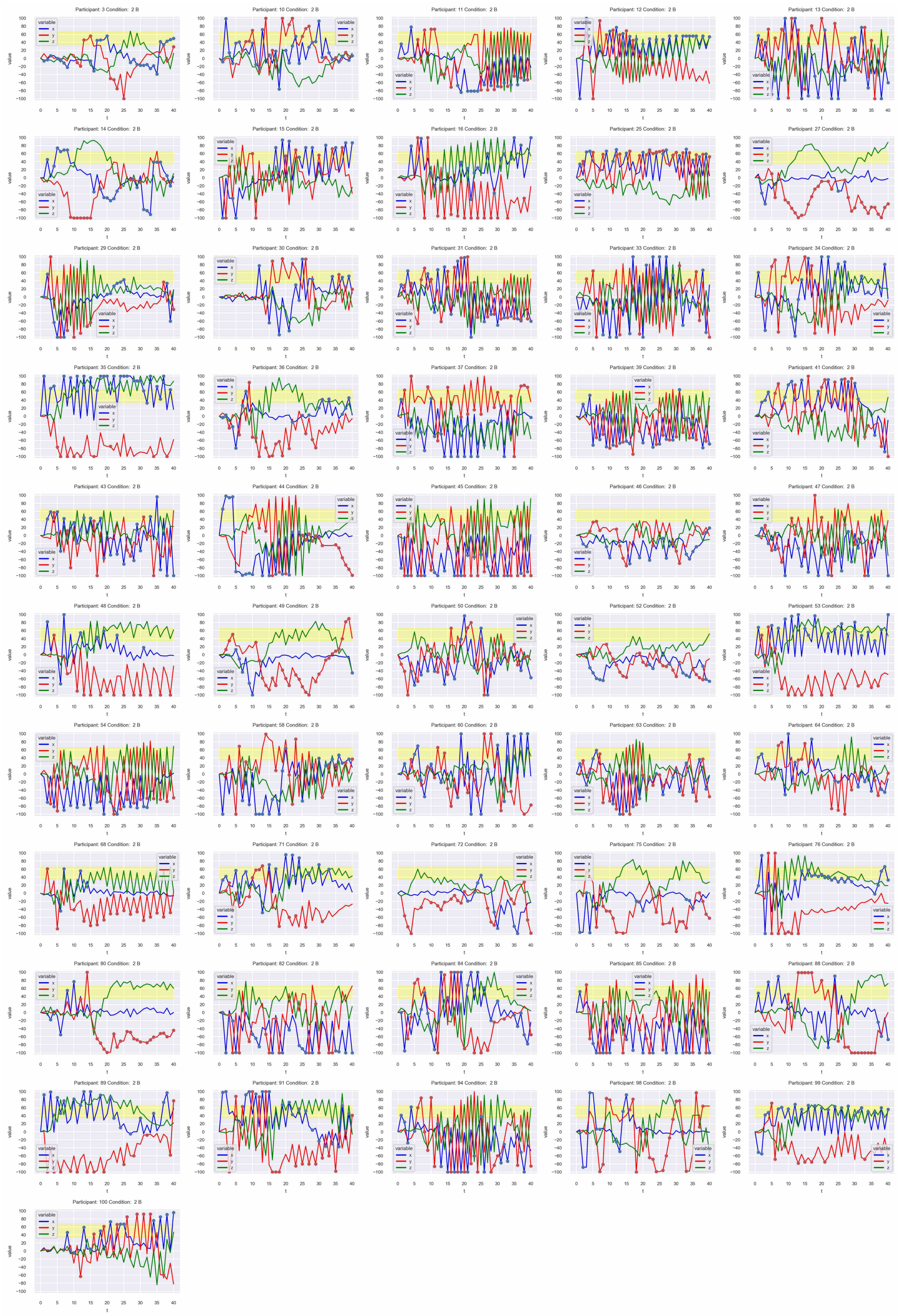

Figure B.4: Participant Response Profiles (Condition = 2 and Group = B). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

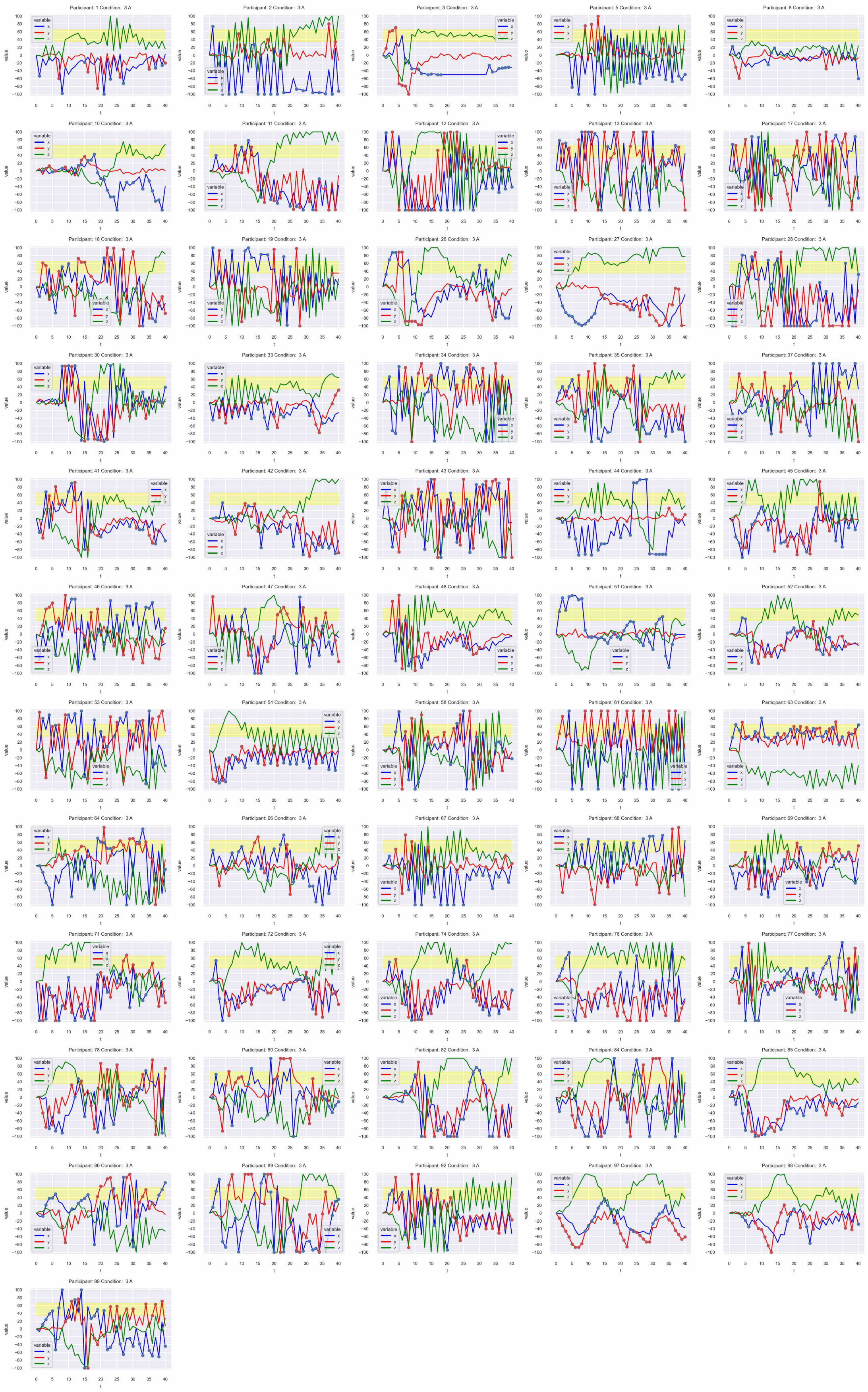

Figure B.5: Participant Response Profiles (Condition = 3 and Group = A). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

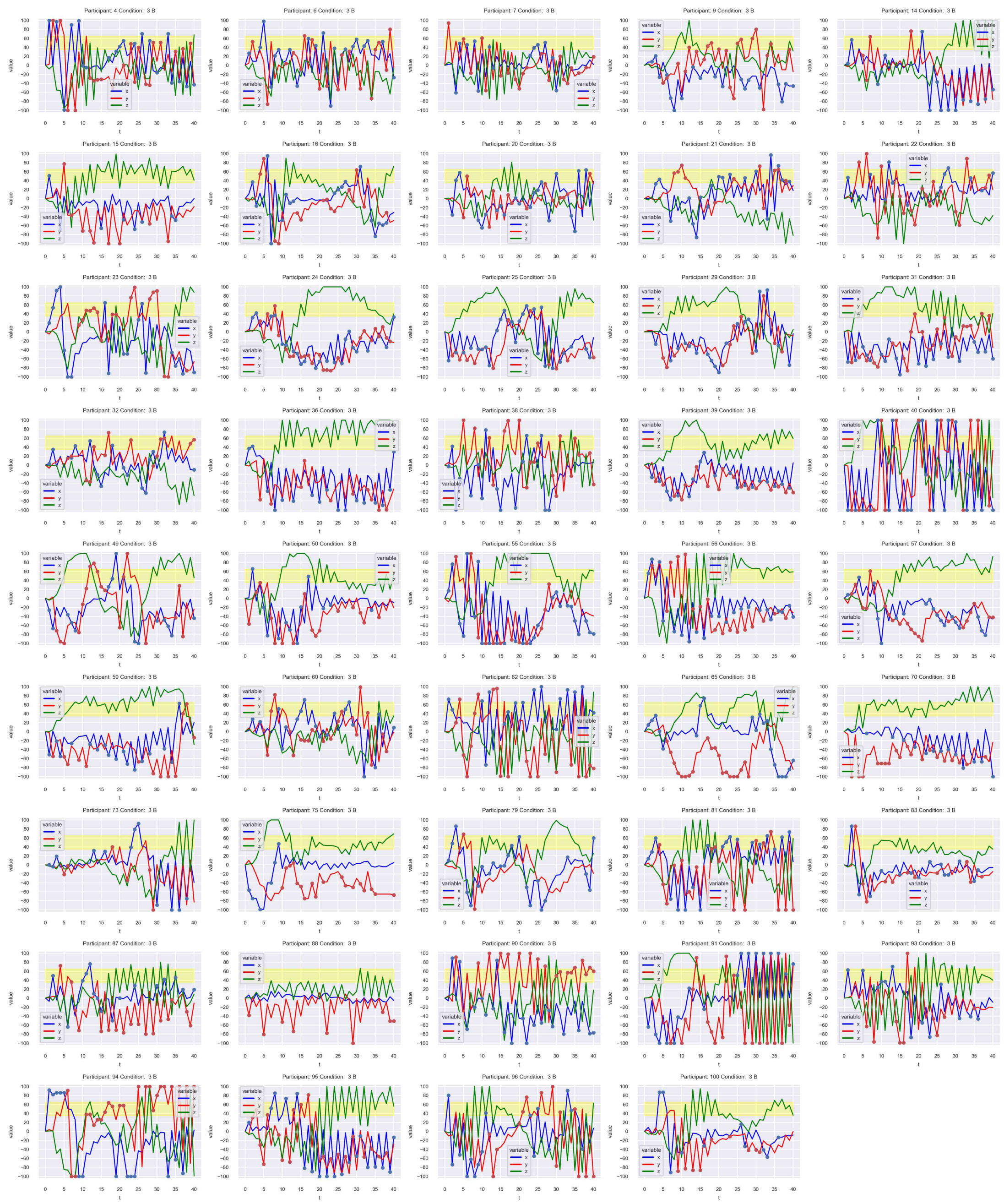

Figure B.6: Participant Response Profiles (Condition = 3 and Group = B). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

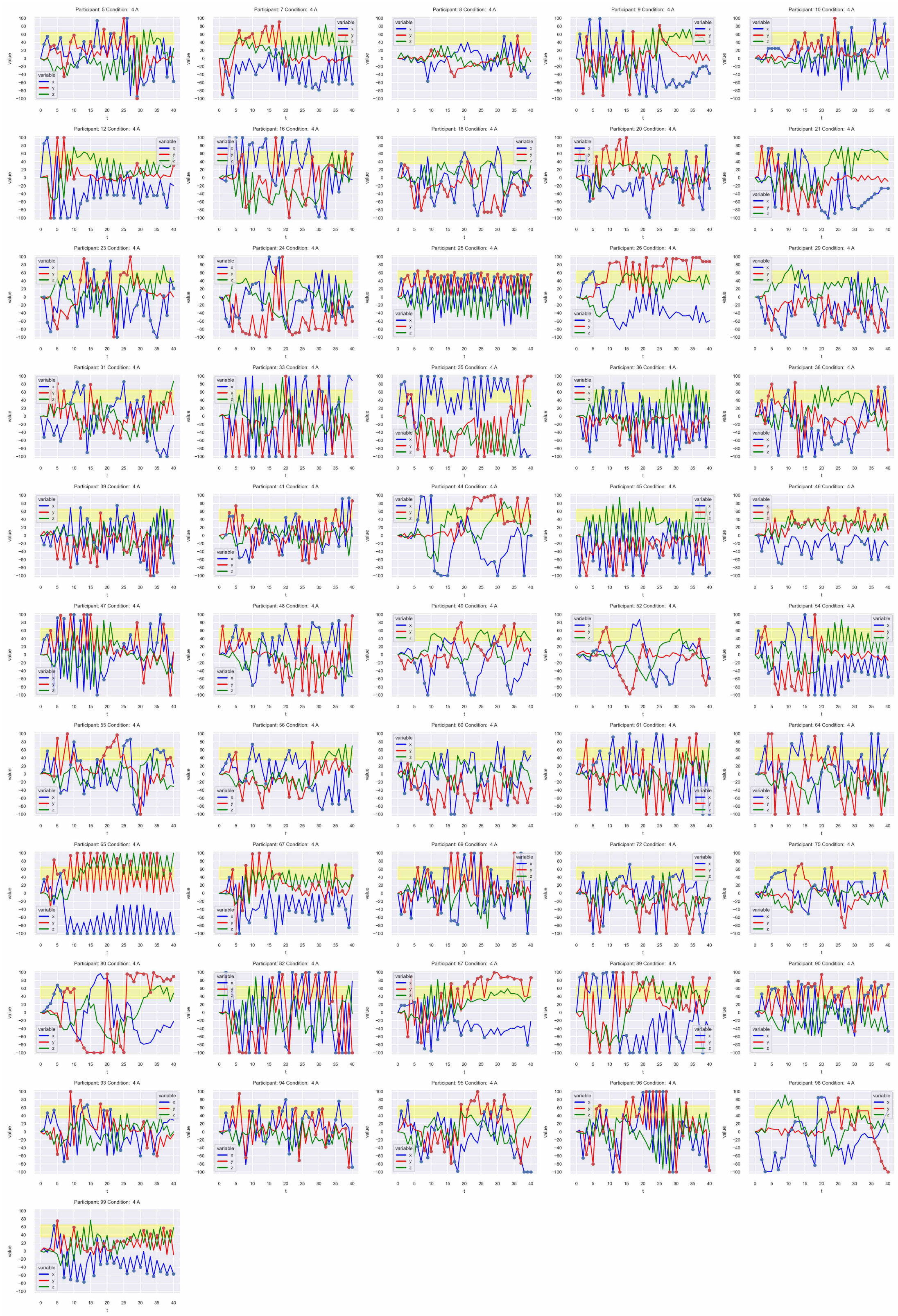

Figure B.7: Participant Response Profiles (Condition = 4 and Group = A). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

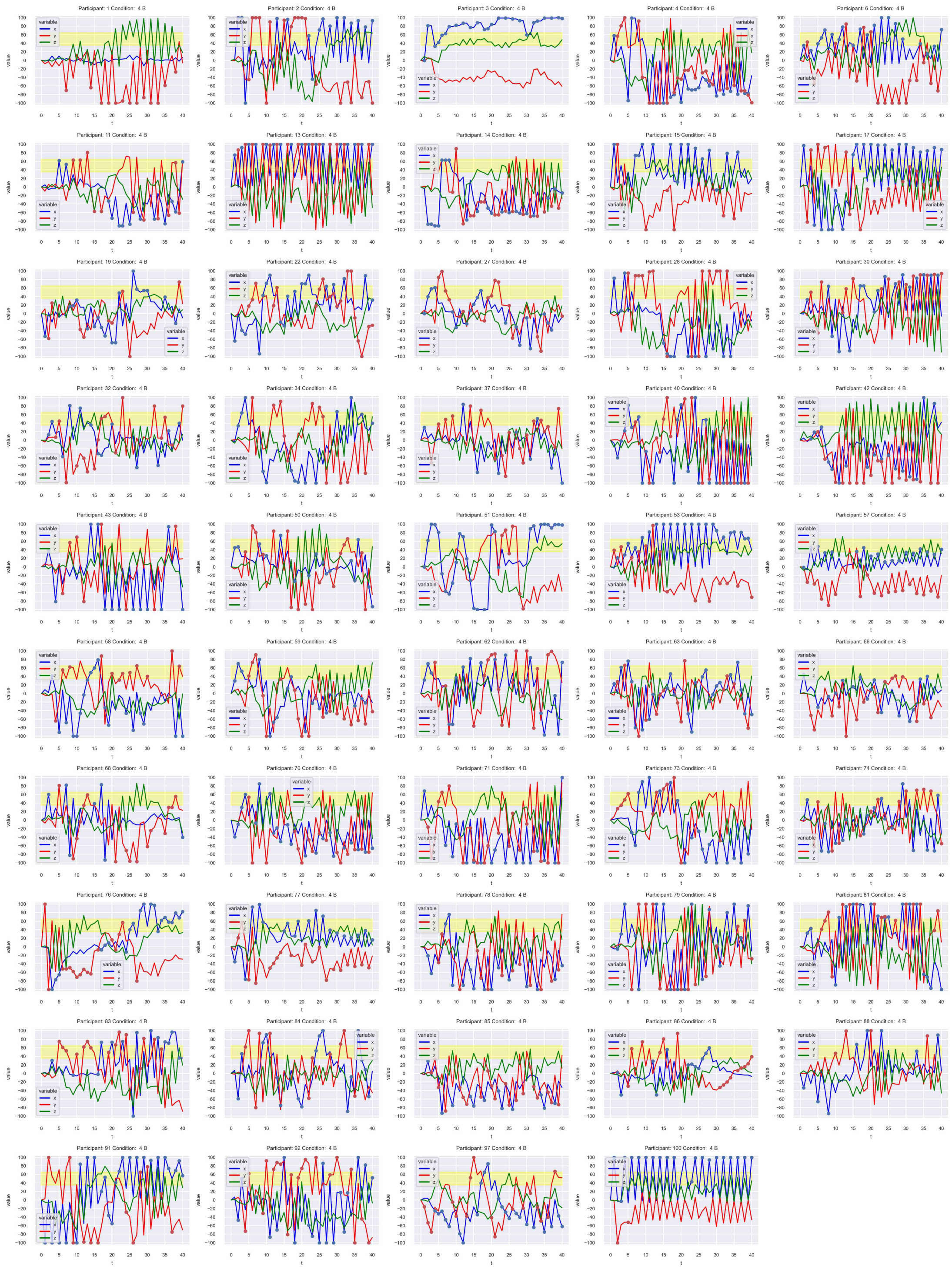

Figure B.8: Participant Response Profiles (Condition = 4 and Group = B). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

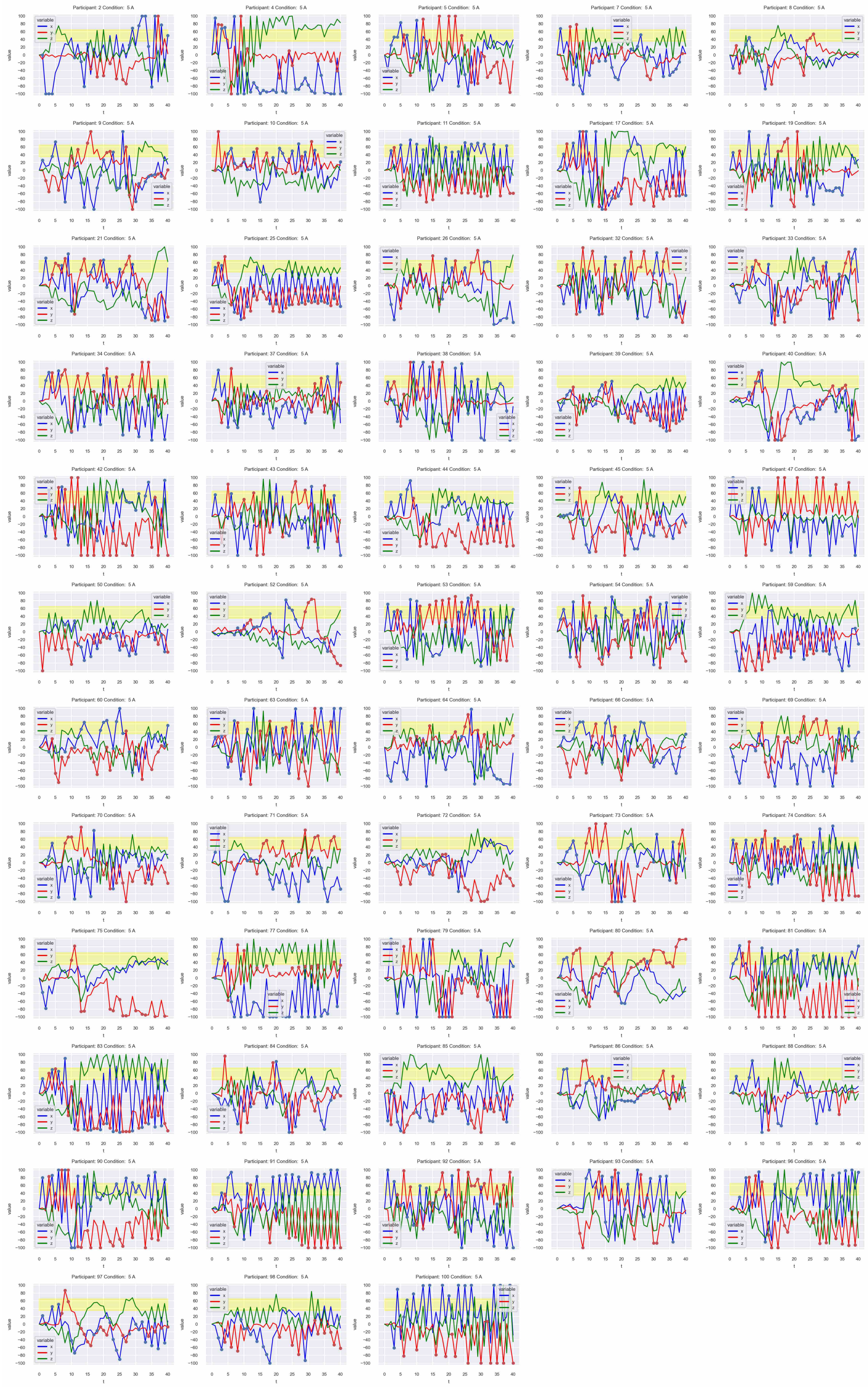

Figure B.9: Participant Response Profiles (Condition = 5 and Group = A). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

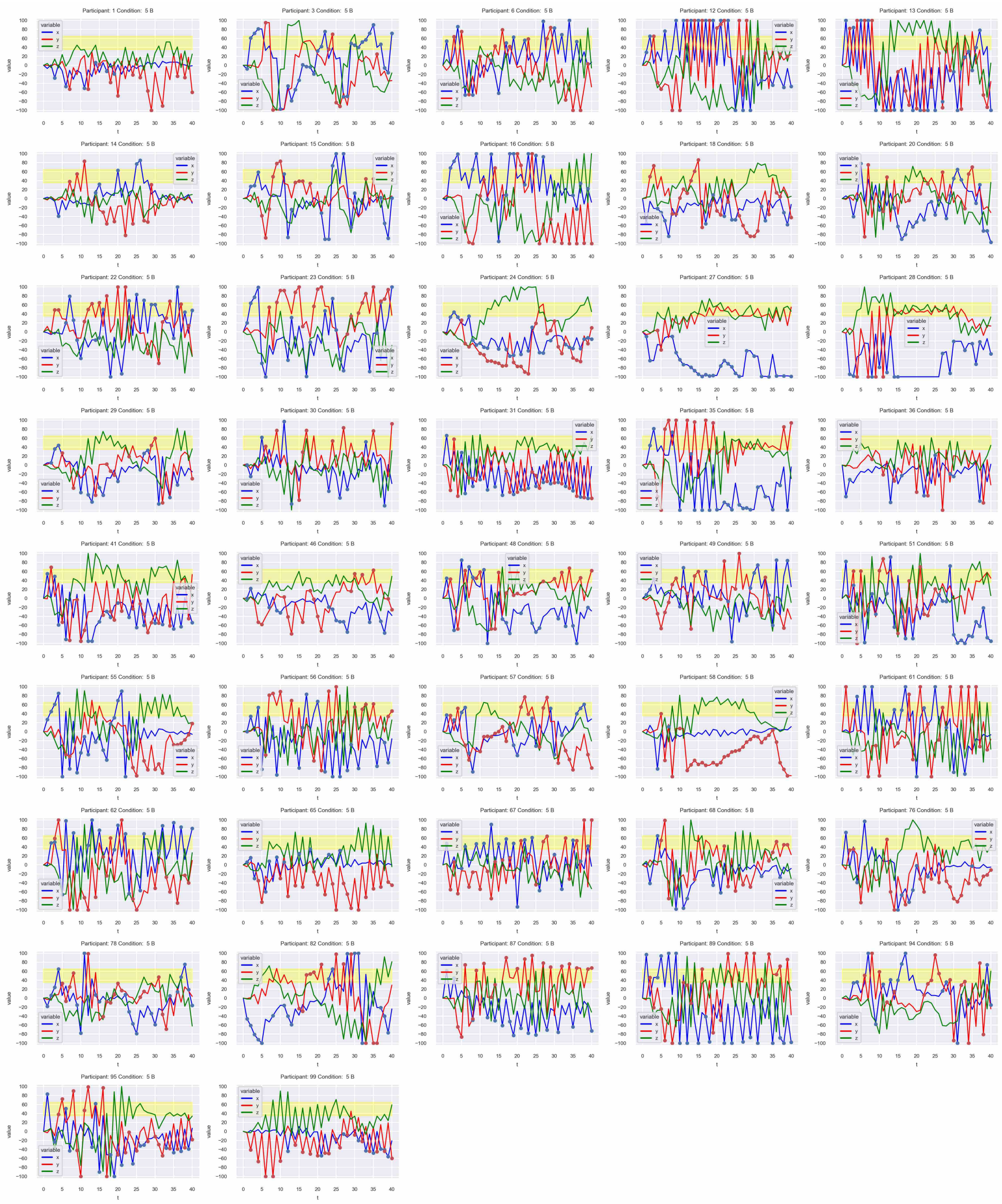

Figure B.10: Participant Response Profiles (Condition = 5 and Group = B). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

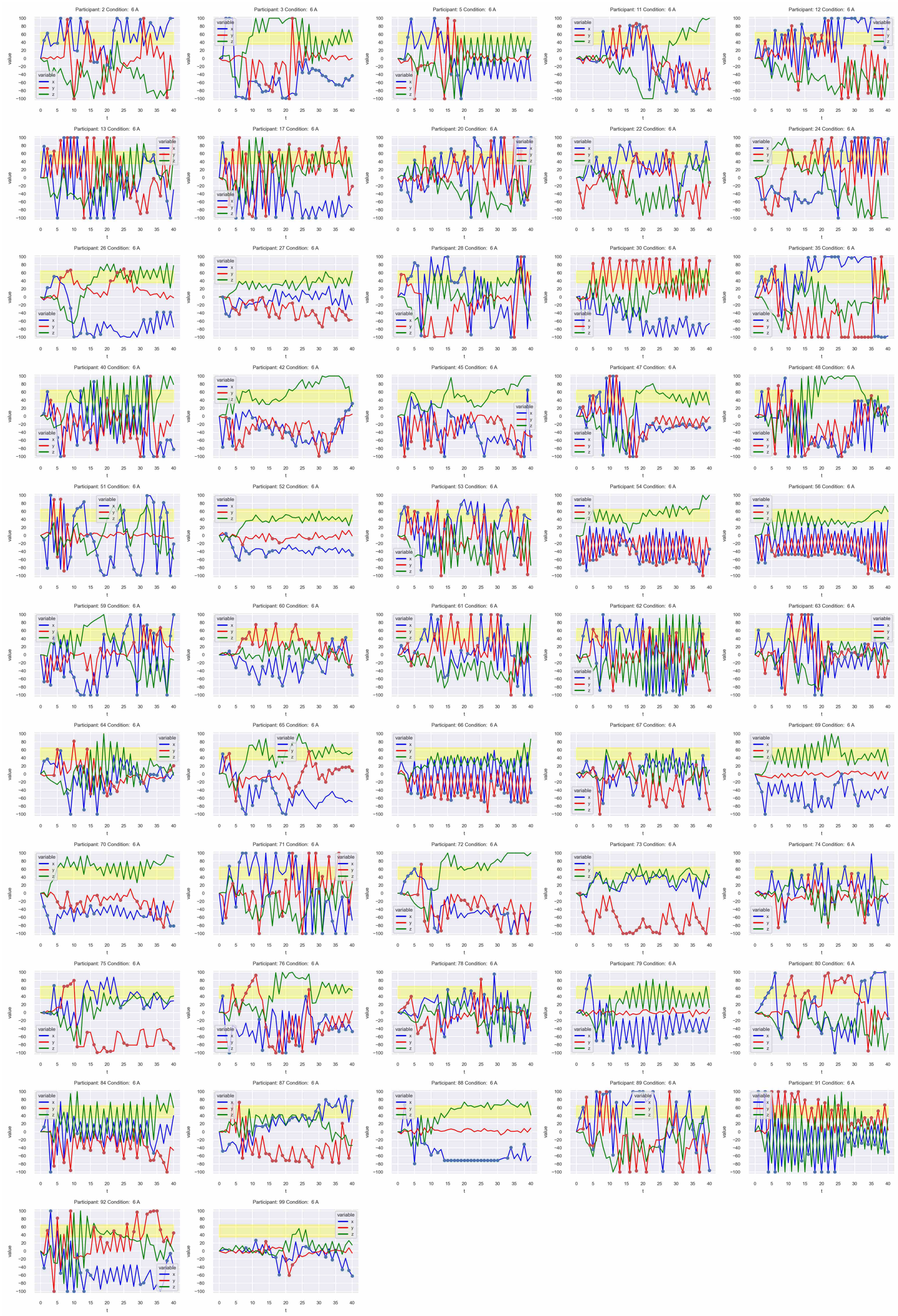

Figure B.11: Participant Response Profiles (Condition = 6 and Group = A). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

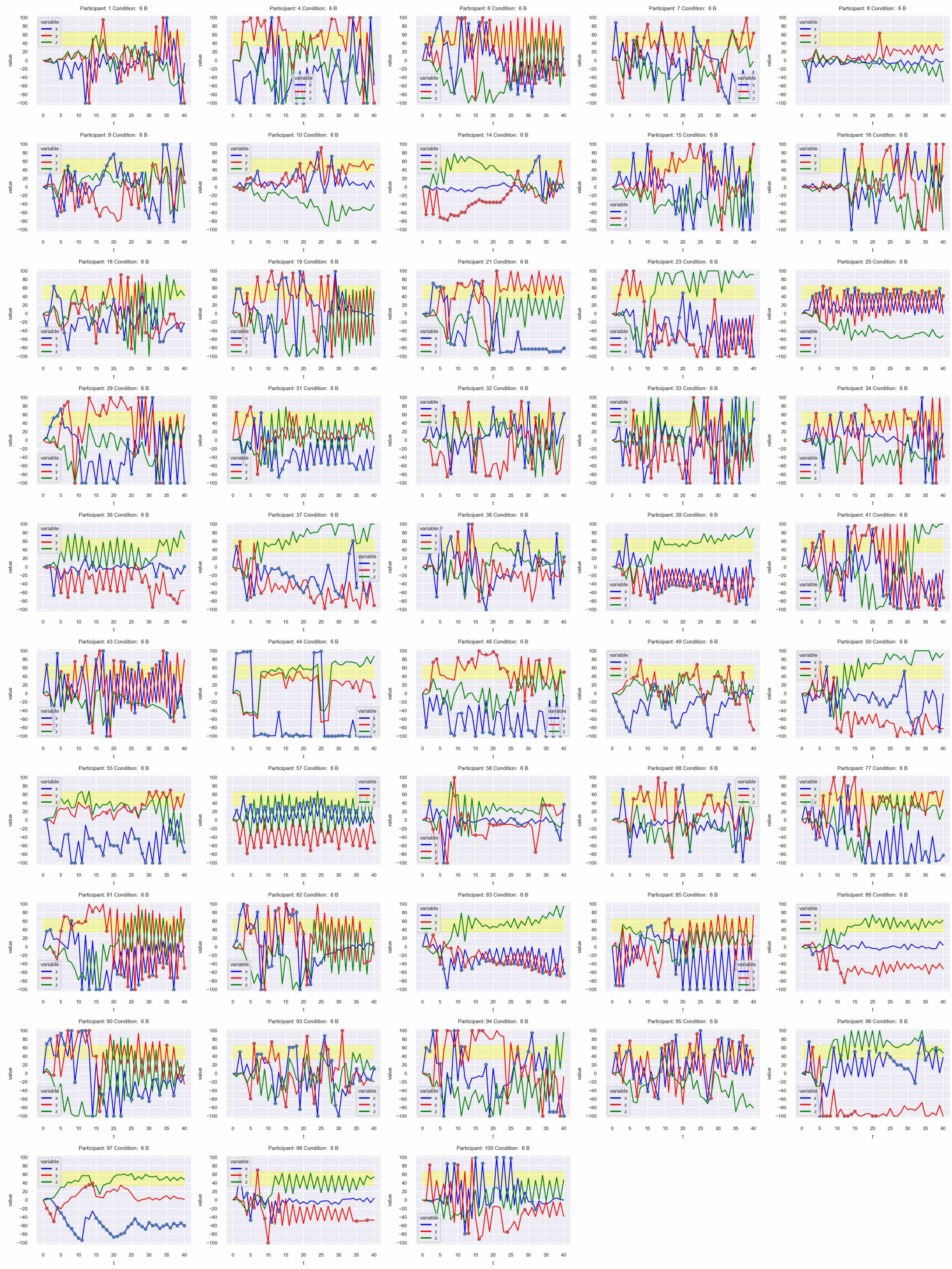

Figure B.12: Participant Response Profiles (Condition = 6 and Group = B). Variables A, B, and C are represented by the colours blue, red, and green respectively. Dots illustrate interventions.

Appendix C

Participant Information Sheet and Consent Form

THE UNIVERSITY *of* EDINBURGH
School of Philosophy, Psychology and Language Sciences

Information sheet for participants

Study title: Adaptive learning and control in dynamic environments
Principal Investigator: Neil Bramley
Researcher collecting data: Paulina Weiss

What is this document? This document explains what kind of study we're doing, outlines your rights, and details the handling of your data. We recommend printing this page for your records.

Nature of the study. You are invited to participate in a study which involves interacting with simple sliders on your screen and answering questions about your observations. Your interactions with the sliders and responses will be recorded. Your session should last for 15 to 20 minutes. You will be given full instructions shortly.

Compensation. For your participation in this study, you will be paid a base rate of £2.50 and a bonus of up to £1.50 depending on your performance.

Risks and benefits. There are no known risks to participation in this study. Other than the payment mentioned, there are no tangible benefits to you, however you will be contributing to our knowledge about causal learning.

Confidentiality and use of data. All the information we collect during the course of the research will be processed in accordance with Data Protection Law. In order to safeguard your privacy, we will never share personal information (like names or dates of birth) with anyone outside the research team. Your data will be referred to by a unique participant number rather than by name. Please note that we will temporarily collect your IP address to prevent repeat participation, however we will never share this information with anyone outside the research team. We will store your IP address using the University of Edinburghs secure encrypted data base and delete it upon completion of the study. The anonymised data collected during this study will be used for research purposes only.

What are my data protection rights? The University of Edinburgh is a Data Controller for the information you provide. You have the right to access information held about you. Your right of access can be exercised in accordance with Data Protection Law. You also have other rights including rights of correction, erasure and objection. For more details, including the right to lodge a complaint with the Information Commissioners Office, please visit www.ico.org.uk. Questions, comments and requests about your personal data can also be sent to the University Data Protection Officer at dpo@ed.ac.uk.

Voluntary participation and right to withdraw. Your participation is voluntary, and you may withdraw from the study at any time and for any reason during the experiment. If you withdraw from the study during data gathering, we will delete your data and there is no penalty or loss of benefits to which you are otherwise entitled. You cannot withdraw from this study after completion of the task because your data will be stored anonymously.

If you have any questions about what you have just read, please feel free to ask, or contact us later. You can contact us by email at s1657612@ed.ac.uk and neil.bramley@ed.ac.uk. This project has been approved by the PPLS Ethics committee. If you have questions or comments regarding your own rights as a participant, they can be contacted at 0131 650 4020 or ppls.ethics@ed.ac.uk.

By clicking the button below, you consent to the following:

1. I agree to participate in this study.
2. I confirm that I have read and understood how my data will be stored and used.
3. I understand that I have the right to terminate this session at any point.

Please enter your Prolific ID:

Figure C.1: Participant information sheet and consent form as presented at the beginning of the study.