Bayesian Analysis of Individual Level Personality Dynamics

Edward Cripps, Robert Wood, Nadin Beckmann, John Lau, Jens F. Beckmann, Sally Wood *

June 22, 2016

*Edward Cripps is Senior Lecturer, Department of Mathematics, University of Western Australia, WA 6009, Australia (E-mail *edward.cripps@uwa.edu.au*). Robert Wood is Professor of Management and Director of Centre for Ethical Leadership, Ormond College, Melbourne Business School and Florey Institute of Neuroscience and Mental Health, University of Melbourne, Carlton 3053 Australia (E-mail *r.wood@cel.edu*). Nadin Beckmann is Senior Lecturer, School of Education, University of Durham, Durham, UK (E-mail *nadin.beckmann@durham.ac.uk*). John Lau is Senior Lecturer, Department of Mathematics, University of Western Australia, WA 6009, Australia (E-mail *john.lau@uwa.edu.au*). Jens Beckmann is Professor of Educational Psychology, School of Education, University of Durham, Durham, UK (E-mail *jens.beckmann@durham.ac.uk*). Sally Wood is Professor of Statistics, Department of Business Analytics, University of Sydney, Darlington 2006, Australia (e-mail *sally.wood@sydney.edu.au.edu*).

Abstract

A Bayesian technique with analyses of within-person processes at the level of the individual is presented. The approach is used to examine whether the patterns of within-person responses on a 12 trial simulation task are consistent with the predictions of ITA theory (Dweck, 1999). ITA theory states that the performance of an individual with an entity theory of ability is more likely to spiral down following a failure experience than the performance of an individual with an incremental theory of ability. This is because entity theorists interpret failure experiences as evidence of a lack of ability which they believe is largely innate and therefore relatively fixed; whilst incremental theorists believe in the malleability of abilities and interpret failure experiences as evidence of more controllable factors such as poor strategy or lack of effort. The results of our analyses support ITA theory at both the within- and between-person levels of analyses and demonstrate the benefits of Bayesian techniques for the analysis of within-person processes. These include more formal specification of the theory and the ability to draw inferences about each individual, which allows for more nuanced interpretations of individuals within a personality category, such as differences in the individual probabilities of spiralling. While Bayesian techniques have many potential advantages for the analyses of processes at the level of the individual, ease of use is not one of them for psychologists trained in traditional frequentist statistical techniques.

1 INTRODUCTION

Psychological reports based on the study of between-person effects often characterize the results as relating to individual level within-person processes. For example, Blackwell, Trzesniewski, and Dweck (2007) describe how, relative to those with an entity or fixed view, individuals with an incremental or developmental view of intelligence “display mastery-oriented strategies (effort escalation or strategy change) versus helplessness strategies (effort withdrawal or strategy perseveration) in the face of setbacks” (Blackwell et al., 2007, p.247). The implication for most readers is that an individual with an incremental view of intelligence will respond to an incident of failure or setback with a mastery oriented strategy, and that an individual with an entity view of intelligence will respond to an incident of failure or setback with a helplessness strategy. The argument that the views, mindsets or beliefs held by individuals shape their reactions to situations, such as failure and setbacks, has been tested for a range of latent variables, including, for example, the ideal versus ought self (Higgins, Roney, Crowe, & Hymes, 1994), learning versus performance goal orientations (Elliott & Dweck, 1988), external versus internal locus of control (Paulhus, 1983) and cultural group processes (Na et al., 2010). In each of these cases, the argument is made that the prior view of each individual influences his or her pattern of responses, but the effects are tested at the group level using aggregate statistics such as means, variances and correlations. Thus statistical inferences regarding between-person differences are used to imply the existence of dynamic within-person processes.

While it is possible that the average pattern of responses observed at the group level will also be observed at the individual level, this cannot be assumed without testing at the individual level (Eysenck & Eysenck, 1985; Borsboom, Mellenbergh, & van Heerden, 2003; Grice, 2015). As noted by Grice (2015, p.1), many relationships observed at the group level do not replicate at the level of the individual, such as the structure of the Big 5 (Grice, Jackson, & McDaniel, 2006; Beckmann, Wood, & Minbashian, 2010) and the Power Law of Learning (Heathcote, Brown, & Mewhort, 2000). While this fact is widely recognized and frequently discussed (e.g., Nezlek, 2001; Schmitz, 2006), a barrier to testing models of psychological processes at the individual level has been an over reliance on the aggregate frequentist statistics of means, variances and correlations that require sample sizes greater than one (Danziger, 1990; Grice, 2015). As a result, the study of individual level processes using, for example, case studies or individual time series to capture the dynamics of within-person processes, such as those described by Blackwell et al. (2007) for entity theorists and incremental theorists, has received relatively little attention until recently.

In more recent times, the collection of individual level time series data with repeated observations of the psychological states and behaviors at multiple time points has been facilitated through the development and application of simulations (R. E. Wood, Beckmann, & Birney, 2009; Beckmann, Wood, Minbashian, & Tabernero, 2012) and experience sampling methods (e.g., Fisher & To, 2012; Minbashian, Wood, & Beckmann, 2010). The analyses of these individual time series has been associated with an increased use of growth curve modelling techniques, including latent curve modelling (LCM; e.g., Goodman, Wood, & Chen, 2011) and growth mixture models (GMM; e.g., Grimm, Ram, & Estabrook, 2010), which combine LCM and finite mixture models to estimate individual trajectories. These methods provide a significant advance in the modelling of dynamic psychological processes in that, in addition to means, variances and correlations they provide estimates of the different trajectories and other features of the pattern of responses over time. However, these are frequentist methods and inference relies on the assumption of asymptotic normality of the sample estimates [1]. While this assumption is generally correct for group level estimates, it is unlikely to be true at the individual level without a large number of observations per individual. As a result, inferences at the individual level from frequentist growth curve modelling techniques are limited to point estimates and do not allow for inferences regarding dynamic within-person processes.

In the current study, we present a Bayesian approach to the modelling of individual level processes using a multiple trial task. Bayesian approaches provide greater flexibility in the modelling of the pattern of within-person processes at the individual level because they are not limited by the assumption of asymptotic normality of the distribution of sample estimates. Given a model to predict the likely observed pattern of individual level outcomes and prior assumptions regarding the parameters that describe the model, Bayesian analyses enable inferences to be made regarding each individual in a sample.

Bayesian analysis offers some advantages for psychologists interested in moving beyond group level tests of between-person differences to study if and how their theories of individual level processes impact on the observed pattern of within-person responses. First is the fact that a Bayesian approach allows for the modelling of individual processes and interpretation of the pattern of observations for each individual in a sample to see if they fit the pattern predicted by the theory. Second, the flexibility of a Bayesian approach requires a priori specification of the processes that generate observations according to the specific theory used to generate the hypotheses, including the predicted pattern of specific values for those observations. The researcher must be able to describe the dynamic model of the processes in mathematical terms, thus requiring greater precision than the prediction of a significant correlation, covariance or mean difference. Third, in the absence of significance tests, Bayesian methods require more detailed examination and explanation of the pattern of results. For example, analyses at the individual level may reveal that most but not all incremental theorists adopt a mastery strategy following failure and that most but not all entity theorists adopt a helplessness strategy. With individual level Bayesian analyses, we are able to determine how many and which individuals in each category respond in a manner that is consistent with the theoretical model and the probability that each individual responds in a manner consistent with their categorization.

[1] The finite sample properties of the estimates in LCM and GMM have not been established.

In the following we will demonstrate how the Bayesian approach can be used to model within-person processes at the level of the individual. We use data from 28 professionals who worked on a complex, dynamic decision-making task and for whom we also collected data about their implicit beliefs about ability.

An Example Study: Implicit Theories of Ability

Two views on intelligence were first described by Carol Dweck as implicit theories of ability (ITA) and later as mindsets (Dweck, 1999), which Dweck labelled as entity and incremental theories. Individuals with an entity theory of ability believe that intelligence is inherent or natural and therefore fixed and not readily subject to change. To the degree that experience and developmental activities make a difference, entity theorists believe it to be the result of pre-existing natural abilities. Individuals with an incremental theory of ability believe that abilities like intelligence are malleable because they are primarily the product of experience, effort and developmental activities. For an incremental theorist, natural abilities are potential to be developed and realized through developmental strategies and effort.

As noted by Blackwell et al. (2007) these two different views of intelligence have been shown to significantly influence how people react to failure and setbacks when learning new tasks (R. E. Wood & Bandura, 1989; Dweck, 1999; Tabernero & Wood, 2010). In her formulation of the ITA model, Dweck (1999) argued that entity theorists who experience failure or setbacks during learning interpret the feedback as evidence of a lack of ability and begin to doubt their capacity to learn the task. If the task is complex enough and requires full use of cognitive resources, this self-doubt interferes with subsequent performance and will lead to a downward spiral. Also, when performing at an acceptable level, entity theorists will stick with the strategy they know and not experiment with new strategies that might expose them to the risk of failure. Thus in the early stages of learning, entity theorists will often lock into a strategy that proves suboptimal as the task unfolds. In contrast, according to Dweck (1999) those classified as incremental theorists are more likely to interpret failure feedback as evidence of a poor strategy or lack of effort. As a result of these attributions to controllable factors, incremental theorists experience less self-doubt and focus on opportunities for improvement by changing their strategy or working harder on subsequent trials, which is more likely to lead to recovery over time.

Thus, the ITA model leads to the prediction that, at an individual level, when performance drops, entity theorists are more likely to spiral further down while incremental theorists are more likely to recover. As a corollary, entity theorists are predicted to learn a task more slowly and have lower performance than incremental theorists, as has been shown at the group level (R. E. Wood & Bandura, 1989; Tabernero & Wood, 2010). As noted above, these aggregated group level results do not directly test the arguments for the differential patterns of individuals' responses to failure by entity and incremental theorists, nor do they demonstrate that the observed group level effects are the product of the predicted dynamics at the individual level. The only conclusion that can be made with confidence in comparisons of the group level learning curves of entity and incremental theorists is that entity theorists, on average, learn at a slower rate than incremental theorists. As well as allowing us to examine group or between person differences in the average rate of performance increase (Question 1), a fuller and more direct analysis of the ITA model at the individual level using Bayesian methods also allows us to examine within person effects (Questions 2 & 3). Our analyses address the following research questions;

1. Do individuals classified as entity theorists increase performance at a slower rate on average than individuals classified as incremental theorists?
2. Following failure what is the likelihood that an individual exhibits spiralling, that is further decreases in performance?
3. Is the probability of spiralling higher for individuals classified as entity theorists than for those classified as incremental theorists?

In addressing these questions we demonstrate features of the Bayesian approach for the analyses of individual level processes and the advantages and disadvantages of that approach. One important advantage of the Bayesian approach for the testing of psychological theories, noted above, is the requirement of specifying how the explanatory mechanisms described in the model will influence the patterns of responses for individuals, plus any assumptions built into the model. Consider research question 2: To answer this question we need to precisely define spiralling behavior in formal mathematical terms and then develop a statistical model to test for its existence. We define spiralling behavior to be a sustained decrease in performance so that individual performance trajectories must be monotonically increasing before the commencement of any spiral and monotonically decreasing afterwards. If individuals' trajectories are assumed to be linear [2] this means that the slopes of these trajectories are positive before and negative after the commencement of a

[2] This is not a necessary assumption, but we use it as a simple example.

spiral. We will show how we incorporate this structure into our model via the prior distribution of the regression coefficients.

The assumption of a prior distribution is sometimes pointed to as a subjective Achilles' heel of Bayesian methods but, in addition to the explicit statement and formal mathematical modelling of the explanatory mechanism and assumptions made, the necessity of specifying a prior distribution allows one to examine the sensitivity of any conclusions to these prior assumptions. For example, in addressing question 3, we ask: How much prior information needs to be imposed in order to conclude that entity theorists are more likely to exhibit spiralling behavior than incremental theorists? We can make inferences about observed differences between entity and incremental theorists using prior beliefs that a difference will occur with a probability ranging from 0% to 100%. Researchers using frequentist statistics are less likely to test the sensitivity of inferences to the assumptions of their models, because the assumptions of asymptotic normality are implicit in the methods so that psychological researchers are often unaware of their existence [3].

Another important feature of Bayesian statistics for analyzing individual level processes is that any event or quantity of interest can be treated as a random variable. In many theories of latent psychological variables that influence individual level processes of learning and performance, the situational event of interest is the experience of failure or a setback. Failures and setbacks are the result of many exogenous forces and can occur at different times for different individuals. This can be modelled as a random variable using Bayesian methods. By way of contrast, psychological experiments based on frequentist methods of inference typically seek to constrain the experience of failure to a single fixed event, a manipulation, and then use aggregate or average group level response to infer individual responses. In Bayesian analyses, the non restrictive assumption of randomness may be applied to a parameter that describes a distribution, such as the mean slope of individual performance trajectories (Question 1), the probability that an individual will start to spiral on a given trial, or it may even be one of a set of statistical models.

These flexible features of the Bayesian approach provide two benefits for the analyses of the individual level processes in response to failure. First is that the trial on which a failure occurs does not have to be fixed but can vary randomly across trials for individuals. Thus, analyses to address questions 2 and 3 do not have to assume that the initial experience of failure is a fixed event that occurs at the same time, or on the same trial, for all individuals in a particular group. But, when

[3] Even when tests for finite samples exist, it is very unusual for psychological researchers to report them.

the experience of failure does occur, be it on trial 3 or trial 10, the responses of entity theorists and incremental theorists will be different. The average performance differences of entity theorists and incremental theorists, even if measured across multiple trials (e.g., R. E. Wood & Bandura, 1989), does not directly test the model proposed by Dweck (1999) and others (e.g., Blackwell et al., 2007) which describe the processes at the individual level when responding to failure events.

Relatedly, Bayesian inference based on the marginal posterior distribution accounts for the joint uncertainty surrounding all unknown parameters. This means that a statement such as "the probability that entity theorists are more likely to exhibit spiralling behavior than incremental theorists is equal to 0.95", accounts for the uncertainty not just in the location of the commencement of the spiral, but also for the uncertainty in the size of individual and group level regression coefficients and error variances. We can therefore be more confident that the effect is real than if we were to plug-in our best guess of the other unknown parameters and compute a *p*-value.

Psychologists interested in analyzing within-person processes at the individual level will also benefit from the fact that Bayesian analyses attach probabilities to each individual's compliance and non compliance with a hypothesis, rather than just reject or accept the hypothesis at the group level. For example, research question 2 will be answered by computing the probability of the two competing models, spiralling or no spiralling, for each individual, based on data available for all individuals. The resulting posterior probability for an individual provides an estimate of the probability that he or she will spiral on future tasks, should we wish to predict the later performance of an individual. For example, we would predict that individual A, for whom the probability of spiralling is equal to 0.99, is much more likely to spiral following failure on a future task than individual B for whom the probability of spiralling is found to equal 0.51.

By way of contrast, the frequentist approach to hypothesis testing would classify both individuals as spirallers and predict that both would spiral following failure on a future task and not differentiate between the probability of each happening. Because the observed pattern of performance for an individual will show that they either spiral or do not spiral, the probabilities of the different models included in the model averaging process must add to 1.0. For example imagine two people, individual A and individual B. For individual A the predictions for spiralling and not spiralling following failure would be weighted by 0.99 and 0.01, respectively. For individual B, the predictions for spiralling and not spiralling following failure would be weighted by 0.51 and 0.49, respectively. Clearly, there would be much greater uncertainty about the prediction for individual

B than for individual A. Frequentist predictions based on model selection ignore the uncertainty associated with the model, and ignoring model uncertainty often leads to *p*-values that overstate the evidence for an effect (Hoeting, Madigan, Raftery, & Volinsky, 1999).

As the number of possible hypotheses or models increases so do the advantages of model averaging over model selection (Raftery & Zheng, 2003). In this paper we average over a very large number of models; for each individual there are 11 possible models, the first specifying no spiral, and within the spiral hypothesis there are 10 sub models, one for each possible location of the trial on which spiralling begins, not allowing spiralling on the last two trials. Therefore, for all 28 individuals the number of possible models is 11^{28}, which is very large indeed. Likelihood based model selection using frequentist procedures, such as AIC or BIC, are not feasible when the number of models under consideration is very large. With such a large number of models we use Markov Chain Monte Carlo (MCMC) methods to stochastically search across the entire model space and predictions are based on a subset of models, rather than a single model, with these predictions weighted by their posterior probability (i.e., the probability of model allocation given the data). Model averaging allows the researcher to ask questions such as "what is the probability that individual j started to exhibit spiralling on trial *i*?"

METHOD

Participants

The participants were 28 managers from various organizations who were attending a three-day executive training program at different times over a year. The twenty-eight participants were all males and had an average age of 34.15 years (SD = 3.23 yrs).

Experimental Task

The experimental task required the participants to manage a computer simulation of a small furniture production and repair workshop containing 5 workers through 12 simulated weeks of business activity (i.e. trials). In this task participants managed the performance of 5 employees by assigning them to each of 5 tasks required to complete a weekly order. The five tasks and the 5 employees remained the same throughout the 12 trials. The challenge for the participants was to learn the optimal match of employees to tasks. The employee performance norm was set at 100 at the start of the task, allowing participants to make judgments about their employees' level of performance (including increase, decrease or otherwise). Trial by trial feedback included the task performance of each of the 5 employees and the overall team performance. The metric for both employees and team performance was hours used as a percentage of budgeted hours for the assigned weekly order, scored so that better performance resulted in higher feedback scores. By using this feedback to test decision options systematically, managers could discover the impact of alternative choices and thereby learn how to increase the organization's performance. Therefore, for each manager there were twelve trials that recorded workgroup performance indicative of managerial ability, which we used as the dependent variable. Further details of the task are described in R. E. Wood and Bailey (1985).

The performance of workers in the simulation had two components; a deterministic component reflecting the consequence of the participant manager's decisions and a random component. The random component was included so that participants could not perfectly predict outcomes, which is a realistic representation of the business world in which managers operate. Note that we chose a dynamic computer simulation that was a novel experience for the participants, for which they had limited expertise and for which they were required to develop new strategies or adapt existing strategies (R. E. Wood & Locke, 1990). New or adapted strategies require greater cognitive effort, have a greater risk of further failure, and require greater persistence in their development and execution than well-known, routine strategies. It is these efforts that are potentially undermined by negative self-evaluations.

Measures

Prior to working on the furniture workshop simulation, participants completed an 8-item measure of their implicit theories of ability (ITA). The 8 ITA items were taken from the measures developed and validated by Dweck and her co-workers (Dweck, 1999) and included four entity type items, such as "People have a certain fixed amount of ability and they cannot do much to change it", and four incremental type items, such as "People can always substantially change their basic skills". All items had a 6-point Likert-type scale ranging from 1 = strongly agree to 6 = strongly disagree. The incremental items were reverse scored and the 8 items were added to create a single scale (*alpha* = .87, *Mean* = 3.41, *SD* = .69), with a higher score indicating a stronger incremental theory and a lower score indicating a stronger entity theory of ability.

A median split was deemed to be an appropriate method of ITA classification as it is the method of categorization for the ITA scale used in Dweck (1999). As a result, the raw data underlying the classifications of participants based on the median split are no longer available; only the coded data has been retained. We acknowledge that using a median split is an increasingly outdated procedure. Nevertheless, we argue that our data are still informative since an individual above the median is more likely to be classified as an incremental theorist than one below the median. Furthermore, the median split provides simpler inferences, although with some loss of granularity, than a continuous variable (e.g., consider the research questions in the Introduction).

Based on a median split of the ITA scores, 14 individuals were classified as entity theorists and 14 classified as incremental theorists. Figure 1 shows the performance of the 28 individuals across 12 trials. Those that are classified as entity theorists are shown in red (*Mean* = 108.42, *SD* = 12.68) and those classified as incremental theorists are shown in blue (*Mean* = 112.1, *SD* = 15.04).

[FIGURE 1 about here.]

BAYESIAN HIERARCHICAL MODEL

We start this section with a hierarchical Bayesian representation of what are commonly called latent curve models (Gelman & Pardoe, 2006; Gelman, 2007) and then demonstrate how the use of prior distributions, together with data augmentation, can be used to extend and tailor these models to answer the questions of interest to psychological researchers.

Consider a series of performance measures on J individuals across T trials. Let $\boldsymbol{Y} = (\boldsymbol{y}_{1.}, \ldots, \boldsymbol{y}_{J.})$, where $\boldsymbol{y}_{j.} = (y_{j1}, \ldots, y_{jT})'$ and y_{jt} is the performance of the j^{th} individual on trial t and denote $f(t)$ to be some function of time. Our purpose in this paper is to demonstrate a number of features of Bayesian methods and therefore we restrict our discussion in the paper to linear functions of time with normally distributed errors. However in Appendix A, we relax these restrictions and consider a nonlinear monotonic function of time and another error distribution.

One possible Bayesian hierarchical model is

$$y_{tj} = \alpha_j + \beta_j t + \varepsilon_{tj}, \quad \varepsilon_{tj} \sim N(0, \sigma^2)$$

$$\alpha_j \sim N(\mu_\alpha, \tau_\alpha^2), \quad \beta_j \sim N(\mu_\beta, \tau_\beta^2), \quad \sigma^2 \sim \mathrm{IG}(a, b) \tag{1.1}$$

where α_j and β_j are the regression coefficients for individual j and the notation $\mathrm{IG}(a, b)$ indicates an inverse gamma distribution with shape and scale parameters a and b respectively. Model (1.1) is a hierarchical one; there are trials within individuals. The model allows individuals to have different regression co-efficients and hence different expected performance trajectories, but the regression co-efficients are restricted to a distribution that depends upon parameters common to all individuals. This distribution is assumed to be normal and the parameters in common are the means, $\boldsymbol{\mu} = (\mu_\alpha, \mu_\beta)$ and variances $\boldsymbol{\tau}^2 = (\tau_\alpha, \tau_\beta)$, of the regression coefficients. These assumptions are not necessary, but are commonly used in Bayesian methods for computational ease, and in frequentist methods because the asymptotic sampling properties of the estimators are known.

The error term in the first line of (1.1) is the within-person variation and $\boldsymbol{\tau}^2$ represents the between individual variation. As $\boldsymbol{\tau}^2 \to (0,0)$ then all individuals have exactly the same expected performance trajectory, while as $\boldsymbol{\tau}^2 \to (\infty,\infty)$ individual expected trajectories have nothing in common with each other and may as well be estimated independently. Clearly the advantage of such a model is that individual trajectories can be estimated based on only a few data points, by "borrowing" information contained in data from other individuals. Note that with only a few data points individual trajectories can only be *estimated*; *inference* surrounding individual trajectories requires the specification of a data generating process such as (1.1), or a large number of data points for each individual.

The model specification is completed by specifying a prior on the hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$. In constructing these priors we use a technique known as Empirical Bayes (Robbins, 1955; Efron, 2005) where the type of prior distribution is specified by the user and then frequentist techniques are used to determine the parameters that describe these prior distributions. For example both μ_α and μ_β are assumed to be independent and normally distributed, centered around the average of the maximum likelihood estimates of the individual regression coefficients, with standard deviations equal to half the range of these quantities. See Appendix C for a full discussion.

Extending and Tailoring the Model

One of the beauties of Bayesian statistics is that, having specified the basic probabilistic data generating process, data augmentation and MCMC techniques can be used to compute the desired characteristic of any posterior distribution. In this section we show how to extend the model in the previous section to answer the research questions described in the introduction.

Using Priors to formulate hypotheses and impose constraints. Research question 1 is relatively straightforward to answer, so we discuss our solution to this before tackling questions 2 and 3. In equation (1.1) we represented a latent curve model as a hierarchical Bayes model in which the unobserved individual regression coefficients, the α's and the β's, are generated from a prior distribution. We now modify this prior to answer specific research questions. There is no reason to suppose, *a priori*, that an individual's ITA classification affects their performance before they have received any performance feedback; as argued above, it is the response to failure feedback and setbacks that differentiates entity and incremental theorists (Dweck, 1999). Therefore, we assume that the prior distribution for the intercept is the same for all individuals, $\alpha_j \sim N(\mu_\alpha, \tau_\alpha^2)$. However in order to answer research question 1 we parameterise our prior for the slope, β_j, to depend upon an individual's ITA classification. Let $\boldsymbol{\mu}_\beta = (\mu_E, \mu_I)'$ and let $z_j = (1,0)$ if individual j is classified as an entity theorists and $z_j = (0,1)$ otherwise. Accordingly $\beta_j \sim N(z_j\boldsymbol{\mu}_\beta, \tau_\beta^2)$, so if an individual is classified as an entity theorist then $\beta_1 \sim N(\mu_E, \tau_\beta^2)$, and if an individual is classified as an incremental theorist, then $\beta_j \sim N(\mu_I, \tau_\beta^2)$. The difference in the mean slopes between the two classifications is given by $\mu_E - \mu_I$ and question 1 is answered by exploring the posterior distribution $p(\mu_E - \mu_I|\mathbf{Y})$; if entity theorists increase performance at a slower rate than incremental theorists then we would expect this distribution to have most of its support less than zero. Note that there is not much practical advantage in using a Bayesian method to answer research question 1. A frequentist approach, such as restricted maximum likelihood estimation, would also suffice and we present a comparison of a frequentist and Bayesian analysis in the Results section.

Answering research question 2 is more complex. As discussed in the introduction, the mean function must be monotonically increasing before and decreasing after the commencement of a spiral. We use the prior distributions of the regression coefficients to enforce these constraints. Suppose the regression function prior to the spiral is given by $\alpha_{1j} + \beta_{1j}t$, where the subscript 1 denotes the function before the spiral. If this function is monotonically increasing then the slope, β_{1j}, must be positive. Similarly suppose the regression function after the spiral is given by $\alpha_{2j} + \beta_{2j}t$, then the slope, β_{2j}, must be negative. In addition these two regression functions must intersect at the commencement of the spiral, which we call the cut point and denote by c_j. To ensure this we need the intercept of the second regression function, α_{2j}, to equal $\alpha_{1j} + c_j(\beta_{1j} - \beta_{2j})$. So we have three constraints (i) $\beta_{1j} > 0$, (ii) $\beta_{2j} < 0$ and (iii) $\alpha_{2j} = \alpha_j + c_j(\beta_{1j} - \beta_{2j})$, all of which can be imposed in a logically consistent manner by the prior. We impose the first and second constraints by assuming that β_{1j} and β_{2j} have normal distributions constrained to be positive and negative respectively. The third constraint is also formulated as a prior distribution, which is that the intercept α_{2j} is equal to $\alpha_{1j} + c_j(\beta_{1j} - \beta_{2j})$ with probability one. Such a distribution function is referred as a Dirac delta function. Note that it is not necessary to think of the prior for α_{2j} as a Dirac delta function, we do so here to show that Bayesian inference is a coherent framework for imposing all model assumptions.

Using Data Augmentation to Model Spiralling. In our response to question 2 we not only want to identify individuals who spiral following failure but we also want to determine the likelihood of spiralling for each individual. That is, we want to be able to say, for example, that "the probability that participant 10 will exhibit spiralling behavior is 0.64". Then, in order to address question 3 we want to determine if the probability of spiralling behavior for each of the 28 participants is related to their categorisation as an entity theorist or an incremental theorist. That is, in addition to modelling behavior at the individual level, researchers also want to understand how group level factors, such as ITA personality classification, affect these individual probabilities of spiralling. In this section we show how data augmentation can answer these questions by facilitating the MCMC scheme that performs the required multidimensional integration needed to estimate the marginal posterior distributions of interest.

To detect spiralling behavior we augment the data with a Bernoulli random variable (*Be*). For each individual we define S_j as

$$S_j = \begin{cases} 1 & \text{if a spiral occurs at any time for individual } j, \\ 0 & \text{otherwise.} \end{cases}$$

If an individual j exhibits spiralling behavior (i.e., $S_j = 1$) we augment the data again with another variable to indicate the point at which the spiral commences, the cut-point, c_j, so that $c_j = t|S_j = 1$ if individual j begins to spiral at time t. The cut-point is a discrete random variable, taking values $1,\ldots,T-2$ and we assume *a priori* that the spiral is equally likely to occur on any trial, therefore $\Pr(c_j = t|S_j = 1) = \frac{1}{T-2}$. Note, under this formulation we do not allow a spiral to begin for the last two trials. The reason for this is to reduce boundary effects and to estimate the regression co-efficient with some precision.

Conditional on S_j and c_j our model for the performance score of individual j on trial t is,

if $S_j = 1$ and $t < c_j$

$$y_{tj} \sim N(\alpha_{1j} + \beta_{1j}t, \sigma^2),$$

if $S_j = 1$ and $t \geq c_j$

$$y_{tj} \sim N(\alpha_{1j} + c_j(\beta_{1j} - \beta_{2j}) + \beta_{2j}t, \sigma^2)$$

with

$$\alpha_{1j} \sim N(\mu_\alpha, \tau_\alpha^2),\ \ \beta_{1j} \sim N_{C_+}(z_j \boldsymbol{\mu}_{\beta_1}, \tau_{\beta_1}^2),\ \beta_{2j} \sim N_{C_-}(z_j \boldsymbol{\mu}_{\beta_2}, \tau_{\beta_2}^2), \tag{1.2}$$

and if $S_j = 0$ then

$$\begin{aligned} y_{tj} &\sim N(\alpha_{1j} + \beta_{1j} t, \sigma^2) \\ \alpha_{1j} \sim N(\mu_\alpha, \tau_\alpha^2),\ \ \beta_{1j} &\sim N_{C_+}(z_j \boldsymbol{\mu}_{\beta_1}, \tau_{\beta_1}^2),\ \ \beta_{2j} \sim \delta(x - a) \end{aligned} \tag{1.3}$$

where $a = 0$.

The notations N_{C_+} and N_{C_-} indicate a normal distribution constrained to be positive and negative respectively. The notation $\delta(x)$ means that $\delta(x) = 1$ if $x = 0$, otherwise $\delta(x) = 0$. So that, in (1.3), $\beta_{2j} = 0$ with probability one.

Note that conditional on an individual spiralling and the location of the cut-point, the estimate of the expected performance trajectory is piecewise linear; $\alpha_1 + \beta_{1j} t$ before the cut point and $\alpha_{1j} + c_j(\beta_{1j} - \beta_{2j}) + \beta_{2j}$ afterwards. However unconditional on these quantities the estimate of the mean performance trajectory is not necessarily piecewise linear. Indeed it will only be piecewise linear if the posterior probabilities of a spiral and corresponding cut-point both equal 1. Figure 2 gives an example of the performance behavior of two individuals. Figure 2, panels (a) and (c) show the estimated posterior mean, $\widehat{E(y_{tj})}$, and posterior probability, $\widehat{\Pr}(c_j|\boldsymbol{Y})$, respectively for individual 20. Panels (b) and (d) are the corresponding plots for individual 28. The fit in panel (b) is close to piece-wise linear, reflecting the fact that the the posterior distribution of c_j is tightly centred around $t = 1$. The nonlinear fit in panel (a) is the result of averaging across several piecewise linear functions, where the averaging is with respect to the posterior distribution of the cut-point.

[FIGURE 2 about here.]

We denote the probability that an individual spirals by $\Pr(S_j = 1) = \pi$, so that $S_j \sim Be(\pi)$ and research question 2 is answered by computing $\Pr(S_j = 1|\boldsymbol{Y})$ for each individual. To answer research question 3, we allow π to depend upon the ITA classification by modelling it as a logistic regression,

$$\pi_j = \frac{\exp(z_j \boldsymbol{\delta})}{1 + \exp(z_j \boldsymbol{\delta})},$$

where $\boldsymbol{\delta} = (\delta_E, \delta_I)$, so that the probability that an entity theorist spirals is $\pi_E = \frac{\exp(\delta_E)}{1+\exp(\delta_E)}$ and the probability that an incremental theorist spirals is $\pi_I = \frac{\exp(\delta_I)}{1+\exp(\delta_I)}$.

We now discuss the prior for $\boldsymbol{\delta}$. If we have no prior belief regarding the probabilities π_E and π_I, other than they must lie between 0 and 1, then the prior on $\boldsymbol{\delta}$ should reflect this. For example in the Appendix we use the prior $\boldsymbol{\delta} \sim N(0, c_\delta \boldsymbol{I}_2)$, where $\boldsymbol{I}_2$ is the 2×2 identity matrix, and show that the choice of $c_\delta = 4$ corresponds approximately to a joint uniform prior. Having established a prior for $\boldsymbol{\delta}$, we answer research question 3 by exploring the posterior distribution $p(\pi_E - \pi_I|\boldsymbol{Y})$. One way of ascertaining the strength of the relationship between the ITA personality type and the propensity to spiral is to see how strong our prior belief must be in order to conclude that there is no relationship. In the results section we show the impact of the value of c_δ has on the posterior density $p(\pi_E - \pi_I|\boldsymbol{Y})$.

Appendix D shows how data augmentation is used to facilitate the MCMC scheme that performs the multidimensional integration needed to estimate the marginal posterior distributions, $p(\mu_E - \mu_I|\boldsymbol{Y})$, $p(\pi_E - \pi_I|\boldsymbol{Y})$.

RESULTS

In this section we present the results for two models; one where the possibility of spiralling is ignored and the other where it is explicitly modelled. Results are categorised as (i) results regarding parameters common to groups of individuals; (ii) results regarding specific individuals; and (iii) results regarding the effect of priors on inference. Model diagnostics, such as residual plots, and simulation results which establish the frequentist properties of the method, are contained in Appendix B.

We present here results for a linear function of time and normal distributed errors. To minimise the risk that any findings are a result of model misspecification consequent upon the choice of a particular function of time, we also obtained results for a logistic growth function, and errors that have a t_ν distribution. The results of these are available in Appendix A and show that the conclusions drawn from the data are unaffected by assumptions regarding these error distributions and functions of time.

Results for parameters common to groups of individuals

First, we examine the results when spiralling is ignored, as described in equation (1.1). Equation (1.1) could also be estimated under the frequentist paradigm and we did so using Restricted Maximum Likelihood (REML), calculated in the `R` package `lme4` (Bates, Mächler, Bolker, &

Walker, 2015). Table 1 reports the results when estimating the parameters common to groups of individuals using both frequentist and Bayesian techniques. The results are very similar. [4]

[TABLE 1 about here.]

A Bayesian analysis of equation (1.1) also allows us to easily estimate $p(\mu_E - \mu_I|\boldsymbol{Y})$, the posterior distribution of the difference between the average rate of learning for entity and incremental theorists. Figure 3 panel (a) is a histogram estimate of this posterior distribution and shows support for research question 1; on average entity theorists learn more slowly than incremental theorists, with probability 0.98. In other words, given the data and prior, the probability that incremental theorists learn at a faster rate is 0.98. Figure 3 panel (a) reports this by showing approximately 0.98 of the mass of $p(\mu_E - \mu_I|\boldsymbol{Y})$ lies below zero.

As noted in the Introduction, when modelling spiralling behavior explicitly in our data, as in equations (1.2) and (1.3), a frequentist analysis is not feasible. We therefore turn our attention to Bayesian analyses only for the rest of the article. Figure 3 panel (b) shows the histogram estimate of $p(\mu_E - \mu_I|\boldsymbol{Y})$ when the existence of spiralling is explicitly modelled. These histograms show that the difference in the learning rate between the two ITA classifications disappears after controlling for the possible existence of spiralling behavior.

[FIGURE 3 about here.]

Figure 4 contains a histogram estimate of the posterior distribution, $p(\pi_E - \pi_I|\boldsymbol{Y})$, and shows that the probability of spiralling is much higher for entity theorists than for incremental theorists, with $p(\pi_E > \pi_I|\boldsymbol{Y}) \approx 0.96$.

[FIGURE 4 about here.]

Individual Level Results

Figure 5 shows the individual posterior mean performance trajectories for entity theorists (red) and incremental theorists (blue), for the model that allows the possibility of spiralling. Panel (a) shows the fit for all individuals. Panel (b) shows the figure for those individuals for whom the probability

[4] We note that, for the frequentist analysis, the sample size may be inadequate for Gaussian approximations to the sampling distributions of estimators and that sampling distributions of estimators of individual level trajectories are not available.

of spiralling was less than 0.5, and panel (c) the figure for individuals for whom the probability of spiralling was greater than 0.5. The three panels of Figure 5 show that while entity theorists are more likely to spiral, not all do. Five out of fourteen did not. Only one out of fourteen incremental theorists exhibited spiralling behavior. Panel (c) also shows that when it is very probable that an individual spirals, the change in that individual's performance trajectory is substantial.

[FIGURE 5 about here.]

Table 2 shows the posterior probability of spiralling for all 28 individuals. A * or * indicates an individual classified as an entity theorist or incremental theorist respectively, for whom the probability of spiralling is greater than 0.5. An estimate of the median value of the point at which the spiral begins, $\hat{c}_j$, is given in the last column. This table shows that the probability of spiralling and the point at which this spiral begins varies between individuals of the same personality classification and demonstrates the need to model behavior at the individual level.

[TABLE 2 about here.]

Effect of Priors on Results

Figure 6 shows the impact that the choice of the prior variance of $\boldsymbol{\delta}$, c_δ, has on the posterior probability $\Pr(\pi_E > \pi_I|\boldsymbol{y})$). Figure 6 shows that the conclusion that entity theorists are more likely to spiral than incremental theorists is largely unchanged in the range $1 < c_\delta < 20$. Indeed the strength of this result can be seen by examining how much prior information needs to be imposed before the result is no longer apparent. From Figure 6 it can be seen that $c_\delta \leq 0.01$ before the $P(\pi_E > \pi_I|\boldsymbol{Y}) \leq 0.5$. In other words we must be 95% certain *a priori* that the probabilities, π_I and π_E, lie in the interval [0.45,0.55], before we would conclude that, on the balance of probabilities, individuals classified as entity theorists are not more likely to spiral than those classified as incremental theorists. For a full discussion of the choice of c_δ see Appendix C.

[FIGURE 6 about here.]

DISCUSSION AND CONCLUSION

In this paper we have presented a Bayesian analysis for the testing of within-person processes at the level of the individual, as well as providing the group level analyses that are usually reported in psychological research using frequentist statistical methods. The contributions and related implications of the reported study can be broken into three categories, which are discussed in turn. First, we discuss the advantages of the Bayesian method for psychologists who wish to study within-person processes at the level of the individual. Second, we discuss the results for the Bayesian analyses of the dynamic model of individual level performance outlined in the ITA model described by Dweck (1999) and the implications for testing other theories of motivation and personality at the individual level. Third, we discuss the functionality of the demands of Bayesian methods for psychologists.

The Bayesian approach provides several advantages over the more commonly used frequentist techniques for psychologists who wish to understand how within-person processes are manifest in the behavior of individuals. First, it allows inference at the individual level even when there are relatively few observations per individual, which is typically the case in longitudinal studies in personality and social psychology. In the current study, there were 12 observations per individual and we were able to test a complex dynamic model as specified by the theory. By way of contrast, if we were to rely on asymptotic arguments that underpin frequentist use of aggregate statistics for inference we would have required many more observations per person and a complex model of the type tested would require a sample of many multiples of that number. Psychological research is expensive and Bayesian methods are more efficient, as well as being more effective in enabling inferences about individuals. This is not an argument for small samples; the cost of obtaining individual level inference is that one must specify a model that generates the data and prior distributions for parameters. Like frequentist methods, Bayesian methods provide more reliable inference with larger samples. Unlike frequentist methods, Bayesian inference is based on the posterior distribution that is calculated using the given observed sample. Of course, in Bayesian statistics a small sample size may mean the prior distribution has a large influence on the posterior distribution. Note, however that one can test the effects of prior specification on the results, as done in this study.

Second, the specification of the prior required by Bayesian methods is a formal mechanism for spelling out the assumptions and prior knowledge of the theory to be tested. This is a discipline that is not required by frequentist approaches but one that will require psychologists to think more critically about the assumptions and current state of knowledge for the theories they employ. Psychologists may not think through the assumptions that underpin the frequentist approaches that they use because there is no formal mechanism or requirement for them to do so. Over time, repeated use of Bayesian methods will begin to lead to common knowledge of priors for different theories and research questions. The current state of knowledge about a relationship can be accumulated on a study-by-study basis. Bayesian methods can also include sensitivity analyses to test for the effects of different priors on the predicted outcomes, as was shown in the results of the current study. Such sensitivity analyses can be used when there is a question about the appropriate prior or when the circumstances suggest that an established prior may not be appropriate due to, for example, challenges to an assumption. The requirement to spell out assumptions and arguments when using Bayesian methods will enable more critical assessments of the cumulative knowledge in psychological research. It will also enable more critical evaluation of populist recommendations, often espoused by consulting firms, that are based on a single study of unknown validity or relevance to the big picture.

Third, Bayesian methods enable researchers to jointly estimate the uncertainty surrounding all parameters. For example, in the current study this enabled us to treat the trial on which an individual experienced their first incident of failure that either did or did not lead to spiralling as a random variable. For psychologists seeking to predict the outcomes of individual processes, the ability to model exogenous factors, such as a performance setback, an action by another person, or some other unexpected event, as random factors, greatly enhances the validity of attempts to model the effects of those events.

This study provided the first test of the individual level performance dynamics of ITA theory. The work of Dweck and colleagues (Dweck, 1999) plus other psychologists who have used ITA theory to develop their hypotheses has been based on an argument that entity theorists respond differently to failure than incremental theorists. In particular, entity theorists are more prone to negative self-evaluation following failure than incremental theorists and these negative self-evaluations are predicted to undermine subsequent performance and lead to spiralling. The data from this study are consistent with the ITA arguments, and further studies are underway to establish the reproducibility of these findings. The results of the current study showed that those identified as entity theorists on a prior independent assessment were more likely on average to exhibit spiralling following an initial failure than those identified as incremental theorists.

We estimated the between-person effect based on the observed within-person response patterns using a bottom up, i.e. individual to group approach, rather than using group-level aggregate statistics to infer the existence of specific response patterns at the level of the individual (top down) as typically done. We also followed recent recommendations to investigate psychological phenomena as a function of time (see Roe, 2008). This enabled us to show not all individuals exhibited the outcomes predicted based on their categorization as either an entity theorist or an incremental theorist, and the onset of the spiralling behavior varied for individuals. These details, which are important for understanding the dynamics and potential limits of the theory are lost in the aggregate statistics of group level analyses. In order to capture these details, we need to model behavior at the individual level, and allow the timing of the commencement of spiralling to vary with individuals.

Approximately two-thirds of the participants classified as entity theorists exhibited spiralling behavior, while the remaining third did not. This is not an uncommon outcome for predictions based on personal characteristics, which are probabilistic and not deterministic. All assessments of the outcomes related to personality characteristics such as ITA have variability and counter indicative results that need to be explained. A further benefit of the Bayesian analyses is that it enables us to identify which of the specific participants categorized as entity theorists did not spiral. Additional knowledge of those individuals and their performance histories can then be explored to see if their deviation from the prediction of the theory are due to problems in the arguments of the theory, boundary conditions of the theory or the fact that they, for whatever reason, did not experience failure during the 12 trails of the simulations. For example, some entity theorists may not have encountered the task conditions that produce failure or they may have discovered effective strategies in the early stages of their task experience. Without the experience of failure, an entity theorist does not experience the self-doubt that can undermine their subsequent performance and behaves like an incremental theorist. Without much larger samples, current frequentist methods cannot identify the performance responses of individuals to specific events. As a result, researchers who use those methods often ignore the variability in predicted outcomes or attribute it to error. Explanations, when offered, are at the group level and refer to characteristics of the sample, the task or the context.

The fact that Bayesian techniques provide individual estimates of the probability of spiralling also has practical implications. For example, if a teacher or counselor was to provide advice to a student identified as an entity theorist, that advice would almost certainly be different for a student with a .95 probability of spiralling following failure in an exam than one whose probability of spiralling is found to equal to 0.51. As noted earlier, the hypothesis selection approach of frequentist statistics would label both as spirallers. The capability of social and personality psychologists to provide more nuanced, individual level analyses of individuals who vary from the mean in their assigned personality category will benefit the clinicians and practitioners who use those categories in their assessments of individuals and resulting interventions. The replication and generalization of the results in further studies will, hopefully, lead to the development of robust priors, this means *a priori* reflections regarding expected effects of tasks, performance profiles and personality constructs. Also, our results might bring spiralling as a general class of response patterns into a more process-orientated focus of attention for different psychological theories that specify differential reactions to success and failure. Another benefit of a Bayesian approach is that it allows updating of estimated probabilities as new evidence comes to hand (rather than abandon old findings and subscribing to new ones, which often is perceived by practitioners as disorientating).

Finally, we turn to the functionality of Bayesian methods for psychologists interested in the study of within-person processes at the individual level. Given the advantages outlined, we might ask why aren't more social and personality psychologists Bayesian? For established scholars whose careers have been built on the understanding and use of frequentist methods, operationalized through standardized statistical packages such as SPSS, AMOS and Minitab, the use of Bayesian methods will present some challenges. Converting the formal mathematical model of the theory into a statistical model requires the use of a range of sampling scheme techniques, such as MCMC, Importance Sampling (IS) and Sequential Monte Carlo (SMC), to efficiently explore the entire model space. The application of these schemes is a non-trivial task and one that often requires mathematical and programming expertise (Browne & Draper, 2006). The flexibility of Bayesian methods to tailor models to answer specific problems, which is one of its strengths, makes the development of off-the-shelf standardized methods problematic. For some researchers who have not had any training in Bayesian statistics these hurdles may seem insurmountable, but not for others. Over many decades, psychology scholars have introduced increasingly sophisticated statistical methods, ranging from factor analyses to growth curve modelling. Depending upon the timing of one's career, scholars have learnt new methods either during their PhD studies or on the job. Over time the introduction of Bayesian statistics training in social sciences will, hopefully, produce a growing body of psychologists who are adept in the flexible application of Bayesian methods and there is evidence that this is a current trend (Andrews & Baguley, 2013).

Of course, not all psychologists interested in the study of dynamic individual level processes need to become experts in Bayesian techniques. Our experience in this research is that collaboration between psychologists and Bayesian statisticians can benefit both disciplines (O'Hagan et al., 2006). Scholars who develop Bayesian methods benefit because often the application of current methods to real problems leads to the development of new methods. Psychologists benefit by being able to construct formal models of their theory and to employ flexible statistical models that provide more direct individual level tests of their theory than less flexible frequentist models. In the current collaboration, the interaction with the Bayesian scholars required clear specification of the arguments and assumptions of the within-person processes in ITA theory and how they would be manifest in an observed pattern of performance over multiple trials, which were then incorporated into the formal model. The specification of the formal model led to great clarity in the specification of the arguments for the ITA theory and the use of highly flexible Bayesian methods enabled the testing of the specified processes at the level of individuals.

Bayesian techniques have the advantage of being more adaptable for specific scientific questions than frequentist techniques. Programs such as R and Winbugs do provide pre-programmed software for some of the standard Bayesian methods used in the analyses of mixture models. However, programmed off-the-shelf software is not yet available for the Bayesian techniques used in the analyses of the complex mixture models required to address specific questions such as those addressed in this manuscript. However, the manuscript provides an explicit description of the MCMC scheme and Matlab code and data can be provided by the authors upon request. The spiralling model may well be one of a general class of models for different psychological theories that specify differential reactions to success and failure, as many social cognitive theories do. For similar, but not identical, applications we argue the collaboration between statisticians and psychologists is necessary to surmount these challenges.

REFERENCES

Andrews, M., & Baguley, T. (2013). Prior approval: The growth of bayesian methods in psychology. *British Journal of Mathematical and Statistical Psychology*, *66*, 1-7.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67(1)*, 1-48.

Beckmann, N., Wood, R. E., & Minbashian, A. (2010). It depends how you look at it: On the relationship between neuroticism and conscientiousness at the within and between-person levels of analysis. *Journal of Research in Personality*, *44(5)*, 593-601.

Beckmann, N., Wood, R. E., Minbashian, A., & Tabernero, C. (2012). Small group learning: Do group members' implicit theories of ability make a difference? *Learning and Individual Differences*, *22(5)*, 624-631.

Blackwell, L., Trzesniewski, K., & Dweck, C. S. (2007). Implicit theories of intelligence achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, *78(1)*, 246-263.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110(2)*, 203-219.

Browne, W. J., & Draper, D. (2006). A comparison of bayesian and likelihood methods for fitting multilevel models. *Bayesian Analysis*, *1(3)*, 1197-1208.

Clyde, M., & George, E. I. (2004). Model uncertainty. *Statistical science*, 81–94.

Danziger, K. (1990). *Constructing the subject*. Cambridge University Press.

Dweck, C. (1999). *Self-theories: Their role in motivation, personality, and development.* Philadelphia, PA: Psychology Press.

Efron, B. (2005). Bayesians, frequentists and scientists. *Journal of the American Statistical Association*, *100*, 1-5.

Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, *54(1)*, 5-12.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach.*

Fisher, C. D., & To, M. L. (2012). Using experience sampling methodology in organizational behavior. *Journal of Organizational Behavior*, *33(7)*, 865-877.

Gelman, A. (2006). Bayesians, frequentists and scientists. *Bayesian Analysis*, *3*, 515-533.

Gelman, A. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.

Gelman, A., & Pardoe, I. (2006). Bayesian measures of explained variance and pooling in multi-level (hierarchical) models. *Bayesian Analysis*, *48*, 241-251.

Goodman, J. S., Wood, R. E., & Chen, C. (2011). Feedback specificity, information processing, and transfer of training. *Organizational Behavior and Human Decision Processes*, *115*, 253-267.

Grice, J. (2015). From means and variances to persons and patterns. *Frontiers in Psychology*, *6(1007)*, 1-12. doi: 10.3389/fpsyg.2015.01007

Grice, J., Jackson, B., & McDaniel, B. (2006). Bridging the idiographic-nomothetic divide: a follow-up study. *Journal of Personality*, *74*, 1191-1218. doi: 0.1111/j.1467-6494.2006.00407.x

Grimm, K. J., Ram, N., & Estabrook, R. (2010). Nonlinear structured growth mixture models in mplus and openmx. *Multivariate Behavioral Research*, *45*, 887–909.

Heathcote, A., Brown, S., & Mewhort, D. (2000). The power law repealed: the case for an exponential law of practice. *Psychonomic Bulletin Review*, *7*, 185-207. doi: 10.3758/BF03212979

Higgins, E. T., Roney, C. J., Crowe, E., & Hymes, C. (1994). Ideal versus ought predilections for approach and avoidance distinct self-regulatory systems. *Journal of Personality and Social Psychology*, *66(2)*, 276.

Hoeting, J., Madigan, D., Raftery, A., & Volinsky, C. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, *14*, 382-417.

Minbashian, A., Wood, R. E., & Beckmann, N. (2010). Task-contingent conscientiousness as a unit of personality at work. *Journal of Applied Psychology.*, *95(5)*, 793-806.

Na, J., Grossmann, I., Varnum, M. E. W., Kitayama, S., Gonzalez, R., & Nisbett, R. E. (2010). Cultural differences are not always reducible to individual differences. *Proceedings of the National Academy of Sciences of the United States of America*, *107(14)*, 6192-6197.

Nezlek, J. (2001). Multilevel random coefficient analyses of event-and interval-contingent data in social and personality psychology research. *Personality and Social Psychology Bulletin*, *27(2)*, 771-785.

O'Hagan, A., Buck, C., Daneshkhah, A., Eiser, J., Garthwaite, P., Jenkinson, D., … Rakow, T. (2006). *Uncertain judgements: Eliciting experts' probabilities*. Wiley Online Library.

Paulhus, D. (1983). Sphere-specific measures of perceived control. *Journal of Personality and Social Psychology*, *44(6)*, 1253.

Raftery, A. E., & Zheng, Y. (2003). Discussion: Performance of bayesian model averaging. *Journal of the American Statistical Association*, *98*, 931-938.

Robbins, B. (1955). An empirical bayes approach to statistics. *Proc. 3rd Berkley Symp. Math.*, *1*, 157–164.

Rodriguez-Yam, G., Davis, R. A., & Scharf, L. (2004). Efficient gibbs sampling of truncated multivariate normal with application to constrained linear regression. *Working Paper*.

Schmitz, B. (2006). Advantages of studying processes in educational research. *Learning and Instruction*, *16(5)*, 433-449.

Tabernero, C., & Wood, R. E. (2010). Implicit theories versus the social construal of ability in self-regulation and performance on a complex task. *Organizational Behavior and Human Decision Processes*, *78(2)*, 104-127.

Wood, R. E., & Bailey, T. (1985). Some unanswered questions about goal effects: A recommended change in research methods. *Australian Journal of Management*, *10*, 61-73.

Wood, R. E., & Bandura, A. (1989). Impact of conceptions of ability on self-regulatory mechanisms and complex decision-making. *Journal of Personality and Social Psychology*, *56*, 407-415.

Wood, R. E., Beckmann, J. F., & Birney, D. (2009). Simulations, learning and real world capabilities. *Education+Training*, *15(5/6)*, 491-510.

Wood, R. E., & Locke, E. A. (1990). Goal-setting and strategy effects on complex tasks. *Research in Organizational Behavior*, *12*, 73-109.

Wood, S. A., Kohn, R., Shively, T., & Jiang, W. (2002). Model selection in spline nonparametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *64*(1), 119–139.

Notes

[1] Edward Cripps is Senior Lecturer, Department of Mathematics, University of Western Australia, WA 6009, Australia (E-mail *edward.cripps@uwa.edu.au*). Robert Wood is Professor of Management and Director of Centre for Ethical Leadership, Ormond College, Melbourne Business School and Florey Institute of Neuroscience and Mental Health, University of Melbourne, Carlton 3053 Australia (E-mail *r.wood@cel.edu*). Nadin Beckmann is Senior Lecturer, School of Education, University of Durham, Durham, UK (E-mail *nadin.beckmann@durham.ac.uk*). John

Lau is Senior Lecturer, Department of Mathematics, University of Western Australia, WA 6009, Australia (E-mail *john.lau@uwa.edu.au*). Jens Beckmann is Professor of Educational Psychology, School of Education, University of Durham, Durham, UK (E-mail *jens.beckmann@durham.ac.uk*). Sally Wood is Professor of Statistics, Department of Business Analytics, University of Sydney, Darlington 2006, Australia (e-mail *sally.wood@sydney.edu.au.edu*).

TABLE 1

Overall performance baseline (μ_α) and performance trajectory (μ_β) as described in equation (1.1) and estimated by a frequentist and Bayesian analysis. Standard errors and posterior standard deviations are in brackets.

	Frequentist $\hat{\mu}_\alpha$	Bayesian $\hat{\mu}_\alpha$	Frequentist $\hat{\mu}_\beta$	Bayesian $\hat{\mu}_\beta$
Incremental Theorist	101.22 (1.77)	101.2 (2.17)	1.67 (0.31)	1.67 (0.38)
Entity Theorist	102.87 (1.88)	102.88 (2.21)	0.3 (0.45)	0.32 (0.55)

TABLE 2

Estimate of posterior means for individual's probability of spiralling, $\hat{\Pr}(S_j = 1|\boldsymbol{Y})$, and posterior medians of the commencement of the spiral, $\hat{c}_j$, for all individuals classified as entity theorists (red) and as incremental theorists (blue) with $f(t) = t$ and $\varepsilon_{jt} \sim N(0, \sigma^2)$. Note that for individual 19, the high probability of spiralling is a result of a low performance score on trial 12. Figure 9 in Appendix A demonstrates how modelling the possibility of large deviations via a t_3 distribution mitigates the impact of outliers.

Posterior Probability of Spiralling					
Incremental Theorists			Entity Theorists		
Individual #	$\hat{\Pr}(S_j = 1\|\boldsymbol{Y})$	$\hat{c}_j$	Individual #	$\hat{\Pr}(S_j = 1\|\boldsymbol{Y})$	$\hat{c}_j$
1	0.11	0	3	0.24	0
2	0.91*	4	5	0.10	0
4	0.09	0	10	0.10	0
6	0.04	0	13	0.05	0
7	0.12	0	14	0.61*	3
8	0.18	0	16	0.33	0
9	0.04	0	18	0.97*	4
11	0.09	0	19	0.99*	9
12	0.22	0	20	0.95*	4
15	0.38	0	21	1.00*	4
17	0.14	0	22	1.00*	3
23	0.02	0	24	0.34	0
25	0.08	0	26	1.00*	3
27	0.12	0	28	0.94*	1
Average	0.18			0.62	

FIGURE 1

Observations on performances over 12 trials for 14 individuals classified as entity theorists (red) and 14 individuals classified as incremental theorists (blue).

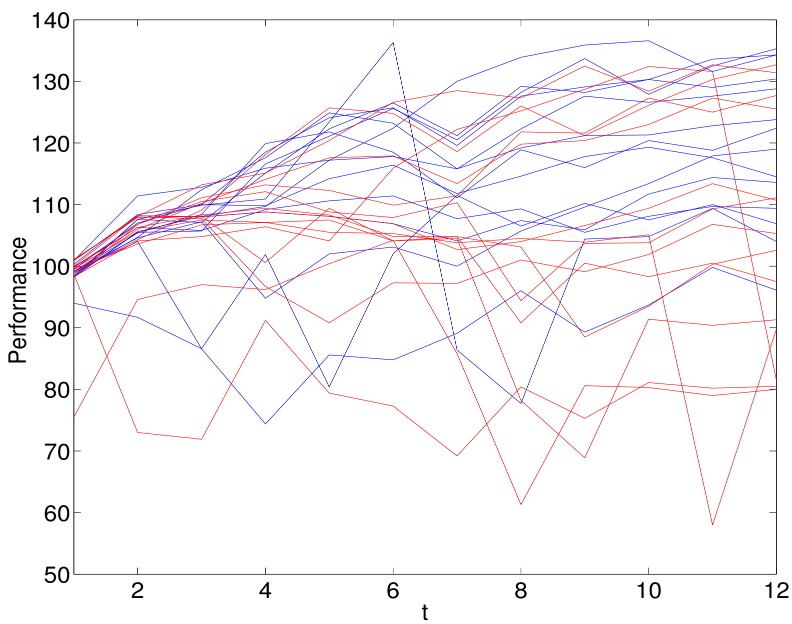


FIGURE 2

Panel (a), shows the data and fitted line for individual 20, who was classified as an entity theorist. The observed data are indicated by '*' and the posterior mean of the regression line is given by the blue line. Panel (c), shows the posterior probability of the commencement of the spiral c_j. Panels (b) and (d) are corresponding plots for individual 28 for was also classified as an entity theorist.

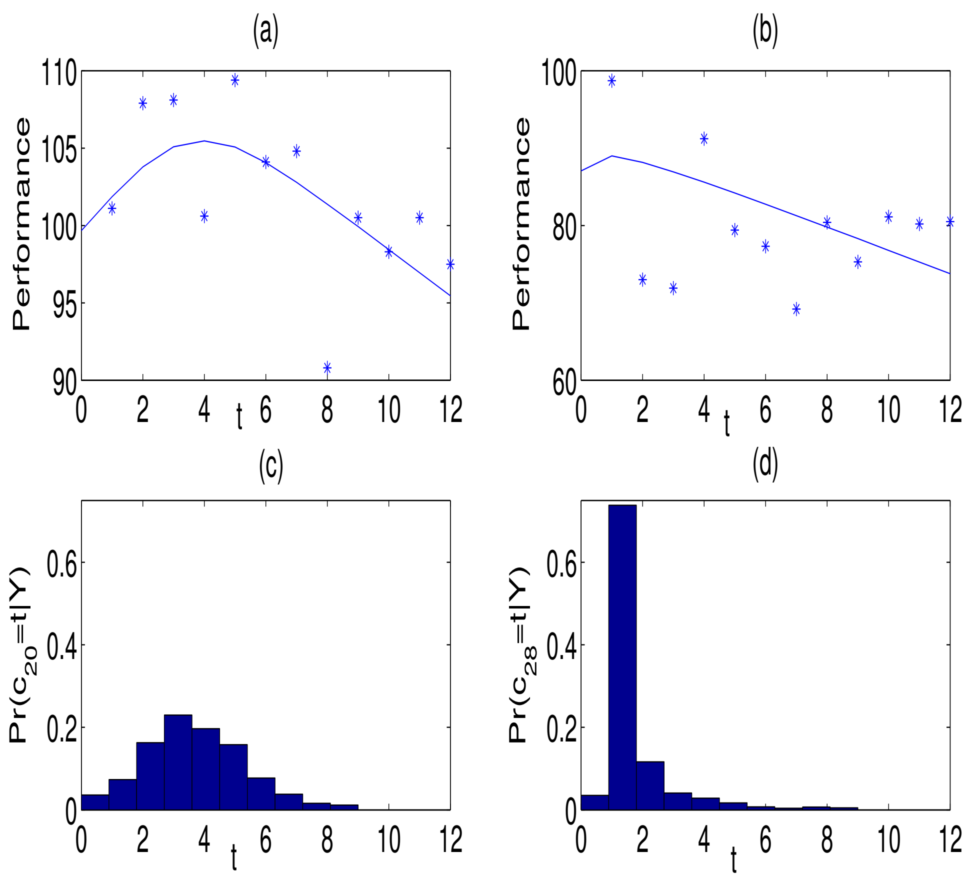


FIGURE 3

Panel (a) reports a histogram estimate of the posterior distribution of $\mu_E - \mu_I$, for the model given by equation (1.1) and $f(t) = t$ and $\varepsilon_{jt} \sim N(0,\sigma^2)$. Panel (b) is a similar plot for the model given by equations (1.2) and (1.3).

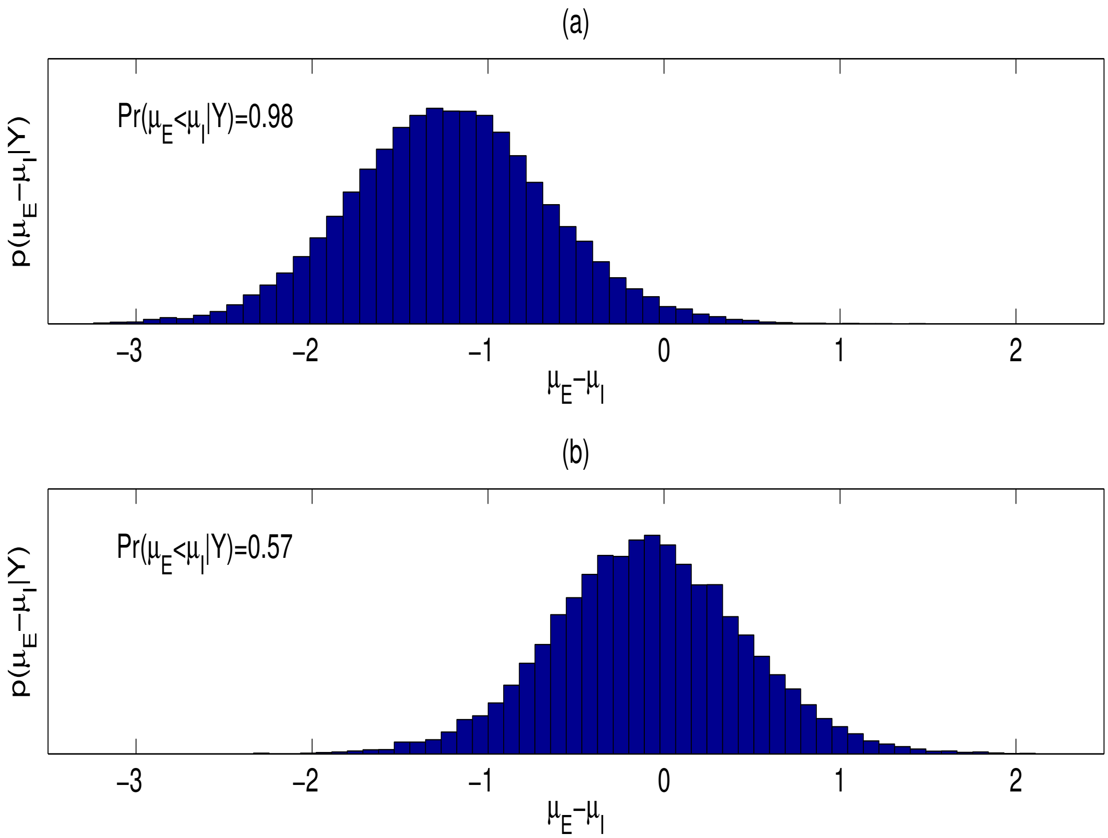


FIGURE 4

Histogram estimate of the difference in the probability of spiralling between entity and incremental theorists, $\pi_E - \pi_I$, for the model given by equations by (1.2) and (1.3) with $f(t) = t$ and $\varepsilon \sim N(0, \sigma^2)$.

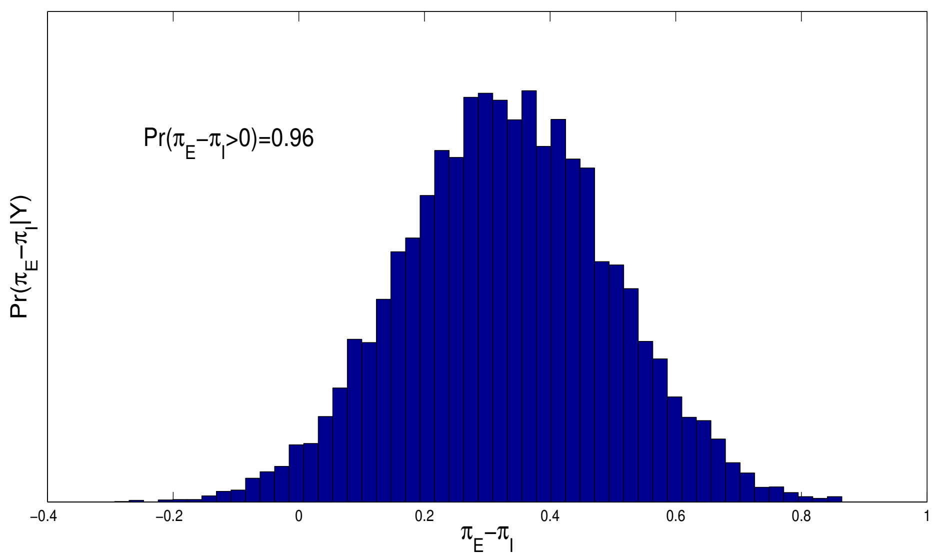


FIGURE 5

Panel (a); Posterior mean of all individual performance curves for entity (red) and incremental (blue) theorists for the model given by equations (1.2) and (1.3), $f(t) = t$ and $\varepsilon_{jt} \sim N(0, \sigma^2)$. Panels (b) and (c) are similar plots for individuals for whom the probability of spiralling is less than 0.5 (panel b) and greater than 0.5 (panel c).

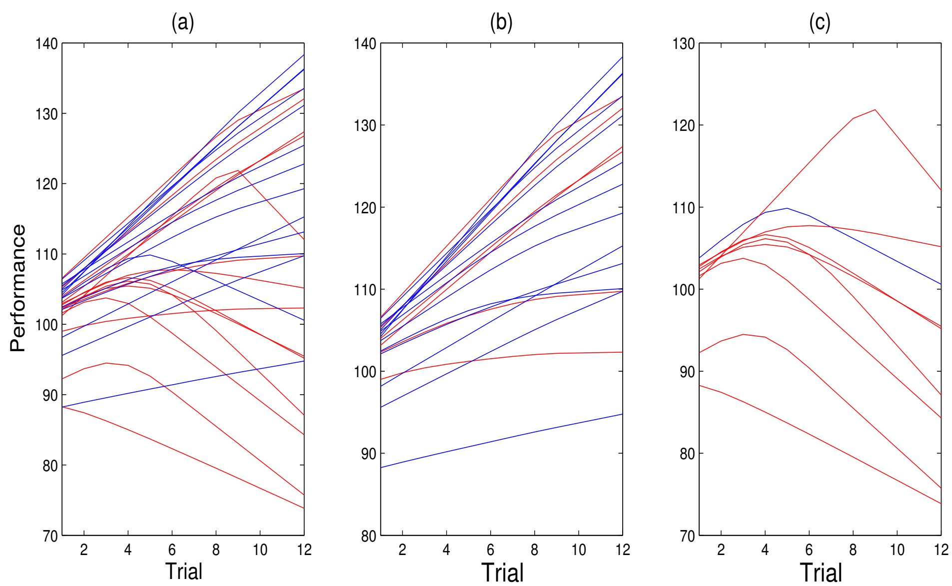
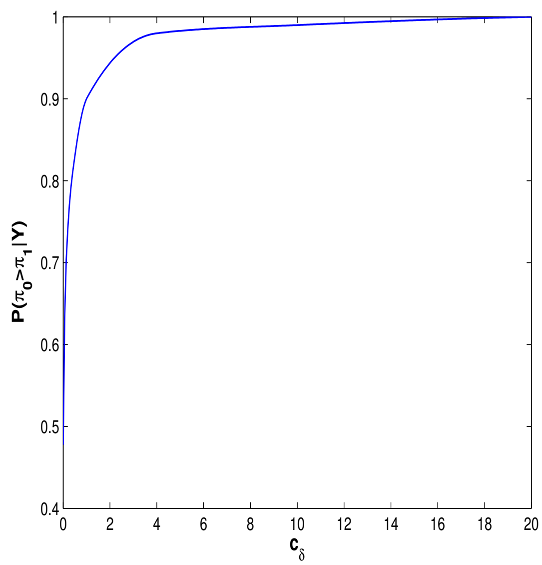


FIGURE 6

The posterior probability that an individual classified as an entity theorist is more likely to spiral than an individual classified as an incremental theorist, as a function of the variance of the prior on $\boldsymbol{\delta}$.



APPENDIX A
Alternate Models

The primary contribution of the statistical method in this article is the Bayesian two level mixture component for random effects models. Modelling this mixture structure as a function of personality type and time permits the estimation of personality group level and also individual level posterior probabilities of (a) the occurence of spiralling behavior and (b) the cut point where spiralling behavior may commence. To stay on point, the main body of the article restricts the discussion to linear mean functions, monotonic either side of the cut point, and Gaussian errors. An advantage of Bayesian methods, coupled with MCMC techniques, is the easy extension to more general models. This allows us to readily fit different models and examine the results, in order to reduce the risk that any findings are a result of model misspecification. We note immediately, in what follows, although some inference at the individual level changes, none of the essential conclusions in the main text are altered, thereby strengthening the support for the ITA.

An equivalent way of writing the two level mixture model in the model development section is for $j = 1,\dots,J$ individuals and $t = 1,\dots,T$ trials

- If $S_j = 0$

$$y_{tj} = \alpha_j + \beta_{1j} f(t) + \varepsilon_{tj}, \quad \varepsilon_{tj} \sim \text{iid} \tag{.4}$$

- If $S_j = 1$ and conditional on $c_j = t^*$

$$y_{tj} = \boldsymbol{x}_t \boldsymbol{\beta}_j + \varepsilon_{tj}, \quad \varepsilon_{tj} \sim \text{iid} \tag{.5}$$

where $\boldsymbol{x}_t = (1, f(t) - (f(t) - f(t^*))^+, (f(t) - f(t^*))^+)$,

$$(f(t) - f(t^*))^+ = \begin{cases} f(t) - f(t^*) & \text{if} \quad f(t) - f(t^*) > 0 \\ 0 & \text{otherwise,} \end{cases}$$

$\boldsymbol{\beta}_j = (\alpha_j, \beta_{1j}, \beta_{2j})'$ and $\varepsilon_{tj} \sim N(0, \sigma^2)$.

Now, write $z_j = (1,0)$ if individual j is an entity theorist, and $z_j = (0,1)$ if individual j is an incremental theorist. We expand the model in 4 ways to allow

1. the observational variance to be parameterized according to personality construct so that incremental and entity theorists have separate variances. That is, for each individual j,

$\sigma_j^2 = z_j(\sigma_E^2, \sigma_I^2)'$. Then if individual j is an entity theorist $\sigma_j^2 = \sigma_E^2$ and if individual j is an incremental theorist $\sigma_j^2 = \sigma_I^2$,

2. the random effects variance parameters to be parametrized according to personality construct. That is $\boldsymbol{\tau}_{\beta_1}^2 = (\tau_{\beta_{1E}}^2, \tau_{\beta_{1I}}^2)'$ and $\boldsymbol{\tau}_{\beta_2}^2 = (\tau_{\beta_{2E}}^2, \tau_{\beta_{2I}}^2)'$.

3. the error structure to have a t_3 distribution, $\varepsilon_{tj} \sim \sigma_j t_3$ to dampen the effects wide tailed error distributions or some extreme values,

4. the learning trajectory to accommodate exponential growth functions where $f(t) = 1 - \exp(-\lambda t)$ depends upon another model parameter, λ. Functions of the form $\alpha + \beta_1(1 - \exp(-\lambda t))$, are often used in the GMM literature because they have the advantage that in addition to being monotonic, an upper and lower limit exists if $\lambda > 0$. If $\beta_1 > 0$ then the lower limit is α and occurs at time $t = 0$, while the upper limit is $\alpha + \beta_1$ and occurs as $t \to \infty$. Conversely if $\beta_1 < 0$, then the upper limit is α, while $\alpha + \beta_1$ is the lower limit. The parameter λ controls the rate at which the function approaches its upper/lower limit. The rate parameters have a random effects structure so each λ_j is generated by a Gaussian distribution, the mean of which depends upon the personality classification of individual j. Also, λ_j is constrained to be positive to ensure that the upper and lower limits exist. We write this as $\lambda_j \sim N_{C_+}(z_j(\mu_{\lambda_E}, \mu_{\lambda_I})', z_j(\tau_{\lambda_E}^2, \tau_{\lambda_I}^2)')$. Then the expected performance score of individual j on trial t conditional on $S_j = 0$ becomes

$$E(y_{tj}) = \alpha_j + \beta_{1j}(1 - \exp\{-\lambda_j t\})$$

and conditional on $S_j = 1$ and $c_j = t^*$

$$E(y_{tj}) = \begin{cases} \alpha_j + \beta_{1j}(1 - \exp\{-\lambda_j t\}) & \text{if} \quad t \le c_j \\ \alpha_j + \beta_{1j}(1 - \exp\{-\lambda_j t^*\}) + \beta_{2j}(\exp\{-\lambda_j t^*\} - \exp\{-\lambda_j t\}) & \text{if} \quad t > c_j. \end{cases}$$

For this choice of function the coefficient, β_{1j}, represents the maximum gain in performance before any possible spiral, not the rate of increase in performance. Also, since the rate which the asymptote is approached is modeled as a random effect, λ_j, the basis function is not common across individuals but rather for individual j is now $f_j(t)$.

Comparison of Results

Figure 7 contains posterior density estimates for the model with $\varepsilon_{tj} \sim \sigma_j t_3$ and $f_j(t) = 1 - \exp\{\lambda_j t\}$.

Panel (a) shows the difference in the probability of spiralling behavior between entity theorists and incremental theorists, $\pi_E - \pi_I$. Panel (a) shows that the probability of spiralling is overwhelmingly higher for entity theorists than for incremental theorists and indeed $\Pr(\pi_E - \pi_I > 0|\boldsymbol{Y}) \approx 0.98$. Panel (b) shows the difference in maximum performance gain before any possible spiral between entity and incremental theorists, $\mu_{\beta_{1E}} - \mu_{\beta_{1I}}$. Panel (b) shows that after accounting for potential spiralling behavior there exists no obvious difference in maximum gain during increasing performance between the two groups: $\Pr(\mu_{\beta_{1E}} - \mu_{\beta_{1I}} < 0|\boldsymbol{Y}) \approx 0.43$.

[FIGURE 7 about here.]

8 shows the individual fitted values when $f_j(t) = 1 - \exp\{\lambda_j t\}$ and $\varepsilon_{tj} \sim \sigma_j t_3$ and supports the results suggested by Figure 7. Figure 8 clearly shows that entity theorists are more likely to exhibit spiralling behavior. Moreover, among those individuals whose probability of spiralling is less than 0.5 (panel (a) Figure 8) there is no obvious difference in performance between entity theorists and incremental theorists. Importantly, Figures 8 and 7 support the broad conclusions of the statistical analysis in the main text regarding the ITA, suggesting model misspecification has not interfered with those aspects of the analysis.

[FIGURE 8 about here.]

Table 3 provides additional insight to differences at the individual level by reporting the posterior probability of spiralling for each individual when $f(t) = t$, $f_j(t) = 1 - \exp\{\lambda_j t\}$ and for $\varepsilon_{tj} \sim N(0, \sigma_j^2)$ and $\varepsilon_{tj} \sim \sigma_j t_3$. This table shows that the probability of spiralling varies between individuals of the same personality classification and demonstrates the need to model behavior at the individual level. Table 3 indicates those individuals who exhibit spiralling behavior – a * or * indicates an individual classified as an entity theorist or incremental theorist respectively, for whom the probability of spiralling is greater than 0.5. The results are fairly consistent, particularly for the posterior median of the cut point, although Table 3 shows different combinations of mean functions and error distributions have a stronger influence inference at the individual level than the group level.

[TABLE 3 about here.]

For instance, consider:

1. Individual 19 has a high probability of spiralling with $\hat{\Pr}(S_{19} = 1|\boldsymbol{Y}) = 0.66$ when $f_j(t) = 1 - \exp\{\lambda_j t\}$ and $\varepsilon_{tj} \sim N(0, \sigma_j^2)$. However this probability drops to 0.07 (with $\hat{c}_{19} = 0$) when $\varepsilon_{tj} \sim \sigma_j t_3$. In main article when $f(t) = t$ and $\varepsilon_{tj} \sim N(0, \sigma^2)$ then $\hat{\Pr}(S_{19} = 1|\boldsymbol{Y}) = 0.99$ and $\hat{c}_{19} = 9$. Figure 9 shows the estimated mean function for the exponential growth model with $\varepsilon_{tj} \sim N(0, \sigma_j^2)$ (dashed line) and with $\varepsilon_{tj} \sim \sigma_j t_3$ (dotted line). This figure shows that extreme observations can have a large impact on the inference regarding individual spiralling behavior. When $\varepsilon_{tj} \sim N(0, \sigma_j^2)$ the extreme observation on trial 12, shown as a '*' , resulted in the method detecting a spiral. However when the possibility of large deviations is explicitly modelled via a t_3 distribution the method does not detect spiralling behavior.

[FIGURE 9 about here.]

2. In the majority of cases the probability that an individual classified as an incremental theorist exhibits spiralling behavior decreases when the mean functions are changed from $f(t) = t$ to $f_j(t) = 1 - \exp\{\lambda_j t\}$. This is because a linear relationship between performance and trial may not be as appropriate as an exponential growth relationship. Perhaps performance increases over time at a decreasing rate and if a linear mean function is used the method occasionally interprets this decrease in the rate of improvement as the beginning of a spiral. Using an exponential growth mean function appears to correct this.

APPENDIX B
Model Diagnostics and Simulations

To check the validity of the model we report residual diagnostics and simulation results. Figure 10 shows that the residuals conform to the model assumption of a $\sigma_j \times t_3$ distribution.

[FIGURE 10 about here.]

Figure 11 displays boxplots of the posterior mean estimates of $\pi_E - \pi_I$ for 3 simulation settings of π_E and π_I, with 50 replications each. The values of π_E and π_I for each setting appear in Table 4. In all settings $\mu_{\beta_{1E}} = \mu_{\beta_{1I}} = 25$; and $\sigma_E^2 = \sigma_I^2 = 15$. These values were chosen because they are close to the posterior mean of the parameters estimated from the data.

[TABLE 4 about here.]

In the first simulation setting the probability of spiralling was zero for both entity theorists and incremental theorists. In the second setting the probability of spiralling was 0.5 for both entity theorists and incremental theorists, while in the third setting the probability of spiralling for entity theorists was set to 0.6, while for incremental theorists it was 0.1. The values of the π's for the third setting were chosen to correspond to the posterior means estimated for the real data. Data were generated from the models given by (.4) and (.5) with $\varepsilon_{tj} \sim \sigma_j \times t_3$.

Figure 11 shows that the median value of the posterior means is very close to the true value for all simulation settings. Additionally when $\pi_E = \pi_I = 0.0$ all the estimated posterior means are tightly centred around zero with an interquartile range (IQR) of [-0.02, 0.01]. However when $\pi_E = \pi_I = 0.5$, there is more variability in the posterior median estimates and the IQR is [-0.19,0.11]. This is to be expected because when spiralling behavior is not present our model detects this, and reduces to a single random effects model. However when spiralling is present, the additional uncertainty surrounding the existence and commencement of spiralling behavior induces additional variability in the parameter estimates.

In simulation setting 3, where all parameters were set to their estimated values for the real data, the boxplots show that the model estimates these parameters well, with the true parameter values very close to the median of the simulation estimates.

[FIGURE 11 about here.]

APPENDIX C
Priors

This paper uses model averaging to make inference regarding the existence of spiralling behavior. The Markov chain Monte Carlo algorithm we constructed is one of varying dimension; if spiralling behavior is not present then there is a single random effects model for performance behavior. If spiralling behavior is present, then performance behavior is described by a mixture of two random effects models, one before the spiral begins and one afterwards. Thus the dimension of the parameter space changes dependent upon which model for individual performance behavior (spiral or no spiral) is generated at each iteration. In model averaging, where the models are nested, the posterior probability of the model with the lowest dimension will be equal to one if improper priors are used, see S. A. Wood, Kohn, Shively, and Jiang (2002) and Clyde and George (2004) for a full discussion. Furthermore even if the dimension of the parameter space is fixed, placing improper priors on parameters in mixture models can result in improper posterior distributions, because there is always the possibility that no observations are allocated to a component in the mixture. For these reasons we place proper priors all parameters.

Prior for $\boldsymbol{\delta}$

Our prior for the probability of exhibiting spiralling behavior is

$$\begin{aligned}\Pr(S_j = 1) &= \frac{\exp(z_j\boldsymbol{\delta})}{1+\exp(z_j\boldsymbol{\delta})} \\ p(\boldsymbol{\delta}) &\sim N(0, c_\delta I_2),\end{aligned}$$

where the parameter c_δ determines the how much the prior shrinks the values of δ_0 and δ_1 toward zero, and hence controls the difference between an entity theorist spiralling and an incremental theorist spiralling, $\pi_E - \pi_I$. If the prior is totally uninformative, i.e. $c_\delta \to \infty$, then we are assuming that the two classifications of personality type have nothing in common regarding the existence of spiralling, and therefore may as well be analysed separately. However as the prior becomes more informative, the probability of spiralling for an individual classified as an entity theorist will approach that of an individual classified as an incremental theorist. In the extreme, if $c_\delta = 0$ then the probability of spiralling for an incremental theorist and an entity theorists will both be equal 0.5 with probability 1. Figure 12 shows the effect of c_δ has on the prior for $\boldsymbol{\pi} = (\pi_E, \pi_I)$. In panel (a), $c_\delta = 1$, in panel (b) $c_\delta = 4$ and in panel(c) $c_\delta = 10$.

[FIGURE 12 about here.]

As this figure shows placing an uninformative prior on $\boldsymbol{\delta}$, by letting $c_\delta \to \infty$ does not result in an uninformative prior for $\boldsymbol{\pi}$. As $c_\delta \to \infty$ the prior weight for π_E and π_I is concentrated on either 1 or 0. Hence choosing a large value for c_δ overstates the difference in the probability of spiralling between entity theorists and incremental theorist. Conversely choosing a small value for c_δ understates the difference in the probability of spiralling between entity theorists and incremental theorist. Choosing $c_\alpha = 4$, approximates a flat prior for π_E and π_I.

Prior for the μ's and τ^2's

We now describe the priors the random effects variances and means parametrized by their personality type. Choosing priors for variance parameters in random effects models can be tricky because of the potential for even weakly informative priors to dominate the information contained in the likelihood. For example using the proper but "non-informative" conjugate inverse gamma prior, IG(a,b), for a variance parameter, where a and b are small, will shrink the posterior distribution of the variance towards zero. For a full discussion of the effect of prior distributions for variance parameters in random effects models see (Gelman, 2006). The potential of the prior to dominate the likelihood is obviously more pronounced if the number of individuals, J, is small. This is a particular problem in this study where the number of individuals who exhibit spiralling behavior can be small. This is a particular problem in this study where the number of individuals who exhibit spiralling behavior can be as small as two or three. To mitigate the potential of the prior to dominate the likelihood we follow (Gelman, 2006) and (Browne & Draper, 2006) place independent uniform priors on the standard deviations of the random effects $\boldsymbol{\tau} \sim U(0, a_\alpha] \times U(0, a_\beta]$. The priors on the hyperparameters are

$$
\begin{aligned}
\mu_\alpha &\sim N(g_\alpha, h_\alpha), \\
\tau^2_\alpha &\sim \mathrm{U}(0, a_\alpha], \\
\boldsymbol{\mu}_{\beta_1} = (\mu_{\beta_{1E}}, \mu_{\beta_{1I}})' &\sim N_{C_+}\left(g_{\beta_1} \times \mathbf{1}_2, h_{\beta_1} \times \boldsymbol{I}_2\right), \\
\boldsymbol{\mu}_{\beta_2} = (\mu_{\beta_{2E}}, \mu_{\beta_{2I}})' &\sim N_{C_-}\left(g_{\beta_2} \times \mathbf{1}_2, h_{\beta_2} \times \boldsymbol{I}_2\right), \\
\boldsymbol{\tau}^2_{\beta_1} = (\tau^2_{\beta_{1E}}, \tau^2_{\beta_{1I}})' &\sim \mathrm{U}(0, a_{\beta_1}] \times \mathrm{U}(0, a_{\beta_1}], \\
\boldsymbol{\tau}^2_{\beta_2} = (\tau^2_{\beta_{2E}}, \tau^2_{\beta_{2I}})' &\sim \mathrm{U}(0, a_{\beta_2}] \times \mathrm{U}(0, a_{\beta_2}],
\end{aligned}
$$

where $\boldsymbol{I}_2$ is the 2×2 identity matrix and $\mathbf{1}_2$ is a vector of ones of length 2. If exponential growth functions are used we have in addition

$$\begin{aligned}\lambda_j &\sim N\left((z_j(\mu_{\lambda_E},\mu_{\lambda_I})', z_j(\tau^2_{\lambda_E},\tau^2_{\lambda_I})'\right)\\(\mu_{\lambda_E},\mu_{\lambda_I}) &\sim N(g_\lambda \times \mathbf{1}_2, h_\lambda \times \boldsymbol{I}_2)\\(\tau^2_{\lambda_E},\tau^2_{\lambda_I}) &\sim \mathrm{U}(0,a_\lambda] \times \mathrm{U}(0,a_\lambda]\end{aligned}$$

and we adopt the following empirical Bayes approach to set the bounds:

1. If $f(t) = t$ denote the maximum likelihood estimate of the mean function coefficients for each individual (when $S_j = 0$) as $(\hat{\alpha}_j, \hat{\beta}_{1j})$ then set

$$\begin{aligned}a_\alpha &= \frac{(\max_j(\hat{\alpha}_j) - \min_j(\hat{\alpha}_j))^2}{4}\\a_{\beta_1} &= \frac{(\max_j(\hat{\beta}_{1j}) - \min_j(\hat{\beta}_{1j}))^2}{4}\\a_{\beta_2} &= a_{\beta_1},\end{aligned}$$

   and $g_\alpha = \sum_{j=1}^J \hat{\alpha}_j/J$, $g_{\beta_1} = \sum_{j=1}^J \hat{\beta}_{1j}/J$, $g_{\beta_2} = -g_{\beta_1}$, $h_\alpha = a_\alpha/\sqrt{J}$, $h_{\beta_1} = a_{\beta_1}/\sqrt{J}$ and $h_{\beta_2} = a_{\beta_2}/\sqrt{J}$.

2. If $f_j(t) = 1 - \exp\{\lambda_j t\}$ denote the maximum likelihood estimate of the mean function coefficients for each individual (when $S_j = 0$) as $(\hat{\alpha}_j, \hat{\beta}_{1j}, \hat{\lambda}_j)$ then set

$$\begin{aligned}a_\alpha &= \frac{(\max_j(\hat{\alpha}_j) - \min_j(\hat{\alpha}_j))^2}{4}\\a_{\beta_1} &= \frac{(\max_j(\hat{\beta}_{1j}) - \min_j(\hat{\beta}_{1j}))^2}{4}\\a_{\beta_2} &= a_{\beta_1},\\a_\lambda &= \frac{(\max_j(\hat{\lambda}_j) - \min_j(\hat{\lambda}_j))^2}{4}\end{aligned}$$

   and $g_\alpha = \sum_{j=1}^J \hat{\alpha}_j/J$, $g_{\beta_1} = \sum_{j=1}^J \hat{\beta}_{1j}/J$, $g_{\beta_2} = -g_{\beta_1}$, $g_\lambda = \sum_{j=1}^J \hat{\lambda}_j/J$, $h_\alpha = a_\alpha/\sqrt{J}$, $h_{\beta_1} = a_{\beta_1}/\sqrt{J}$ and $h_{\beta_2} = a_{\beta_2}/\sqrt{J}$ and $h_\lambda = a_\lambda/\sqrt{J}$.

Prior for σ^2

We set an uninformative uniform prior for the observational variances contained in $\boldsymbol{\sigma}^2$. That is, $p(\sigma_E) \sim U(0,k]$ and $\sigma_I \sim U(0,k]$ for some large non-negative constant k.

APPENDIX D
Sampling Scheme

Write $\boldsymbol{S} = (S_1, S_2, \ldots, S_J)$, $\boldsymbol{C} = (c_1, c_2, \ldots, c_J)$ and for the case when $S_j = 0$

$$\boldsymbol{X}_{j|0} = \begin{bmatrix} 1 & f(1) \\ 1 & f(2) \\ \vdots & \vdots \\ 1 & f(T) \end{bmatrix} \quad \text{and} \quad \boldsymbol{b}_{j|0} = (\alpha_j, \beta_{1j})'$$

and for the case when $S_j = 1$ and conditioned on $c_j = t^*$

$$\boldsymbol{X}_{j|1} = \begin{bmatrix} 1 & f(1) - (f(1) - f(t^*))^+ & (f(1) - f(t^*))^+ \\ 1 & f(2) - (f(2) - f(t^*))^+ & (f(2) - f(c_t^*))^+ \\ \vdots & \vdots & \vdots \\ 1 & f(T) - (f(T) - f(t^*))^+ & (f(T) - f(t^*))^+ \end{bmatrix} \quad \text{and} \quad \boldsymbol{b}_{j|1} = (\alpha_j, \beta_{1j}, \beta_{2j})'.$$

Also, write $\boldsymbol{b}_0 = \{\boldsymbol{b}_{j|0} : S_j = 0\}$, $\boldsymbol{b}_1 = \{b_{j|1} : S_j = 1\}$ and $\boldsymbol{B} = \{\boldsymbol{\alpha}, \boldsymbol{b}_1\}$, $\boldsymbol{\Theta} = (\Theta_0, \Theta_1)$, $\Theta_0 = \{\mu_\alpha, \tau^2_\alpha, \mu_{\beta_{1E}}, \mu_{\beta_{1I}}, \tau^2_{\beta_{1E}}, \tau^2_{\beta_{1I}}\} = \{\mu_\alpha, \tau^2_\alpha, \boldsymbol{\mu}_{\beta_1}, \boldsymbol{\tau}^2_{\beta_1}\}$,
$\Theta_1 = \{\mu_\alpha, \tau^2_\alpha, \mu_{\beta_{1E}}, \mu_{\beta_{1I}}, \tau^2_{\beta_{1E}}, \tau^2_{\beta_{1I}}, \mu_{\beta_{2E}}, \mu_{\beta_{2I}}, \tau^2_{\beta_{2E}}, \tau^2_{\beta_{2I}}\} = \{\mu_\alpha, \tau^2_\alpha, \boldsymbol{\mu}_{\beta_1}, \boldsymbol{\tau}^2_{\beta_1}, \boldsymbol{\mu}_{\beta_2}, \boldsymbol{\tau}^2_{\beta_2}\}$. Finally, to implement the MCMC scheme when the ε_{tj}'s have a scaled t_3 distribution, define $\varepsilon_{tj} = e_{tj}\sqrt{3/(\kappa_{tj})}$, where $\kappa_{tj} \sim \chi^2_3$ and $e_{tj} \sim N(0, \sigma^2)$. Then conditional on κ_{tj} the distribution of $\varepsilon_{tj}|\kappa_{tj}$ is $N(0, \omega_{tj})$ where $\omega_{tj} = \sigma^2 3/\kappa_{tj}$, and ω_{tj} is the t^{th}, diagonal element of a diagonal matrix Ω_j. Finally, write $\boldsymbol{\Omega} = \{\Omega_j : j = 1, 2, \ldots, J\}$.

The sampling scheme is then

1. Sample $\boldsymbol{S}$.

$$p(\boldsymbol{S}|\boldsymbol{Y}, \boldsymbol{\Theta}, \boldsymbol{\Omega}) = \prod_{j=1}^{J} p(S_j|\boldsymbol{y}_j, \Omega_j, \boldsymbol{\Theta})$$

where

$$p(S_j = 1|\boldsymbol{y}_j, \Omega_j, \Theta_1) = \frac{p(\boldsymbol{y}_j|\Omega_j, \Theta_1, S_j = 1)P(S_j = 1)}{p(\boldsymbol{y}_j|\Omega_j, \Theta_1, S_j = 1)P(S_j = 1) + p(\boldsymbol{y}_j|\Omega_j, \Theta_0, S_j = 0)P(S_j = 0)}$$

and

$$
\begin{aligned}
p(\boldsymbol{y}_j|\Omega_j,\Theta_1,S_j=1) &= \frac{\sum_{t=1}^{T-2}}{T-2}\times \\
&\int_{\mathbb{R}\times C_+\times C_-} p(\boldsymbol{y}_j|S_j{=}1,\Omega_j,\Theta_1,c_j{=}t,\boldsymbol{b}_{j|1})p(\boldsymbol{b}_{j|1}|\Theta_1,S_j{=}1)d\boldsymbol{b}_{j|1}\Pr(c_j{=}t|S_j{=}1) \\
p(\boldsymbol{y}_j|\Omega_j,\Theta_0,S_j=0) &= \int_{\mathbb{R}\times C_+} p(\boldsymbol{y}_j|S_j{=}0,\Omega_j,\Theta_0,\boldsymbol{b}_{j|0})p(\boldsymbol{b}_{j|0}|\Theta_0)d\boldsymbol{b}_{j|0} \qquad (.6)
\end{aligned}
$$

The integrals in (.6) are equal to

(a)

$$
\begin{aligned}
p(\boldsymbol{y}_j|\Omega_j,\Theta_0,S_j=0) &= \frac{|\boldsymbol{T}^*_{j|0}|^{1/2}}{(2\pi)^{T/2}|\boldsymbol{T}_{j|0}|^{1/2}|\Omega_j|^{1/2}} \\
&\times\exp\left\{-\frac{1}{2}\left(\boldsymbol{y}_j'\Omega_j^{-1}\boldsymbol{y}_j+\boldsymbol{M}'_{j|0}\boldsymbol{T}^{-1}_{j|0}\boldsymbol{M}_{j|0}-\boldsymbol{M}^{*\prime}_{j|0}\boldsymbol{T}^{*-1}_{j|0}\boldsymbol{M}^*_{j|0}\right)\right\} \\
&\times\frac{1-\Phi\left((\infty,0)'|\boldsymbol{M}^*_{j|0},\boldsymbol{T}^*_{j|0}\right)}{1-\Phi\left((\infty,0)'|\boldsymbol{M}_{j|0},\boldsymbol{T}_{j|0}\right)}
\end{aligned}
$$

where

$$
\boldsymbol{T}_{j|0}=\begin{bmatrix}\tau^2_\alpha & 0\\ 0 & z_j\boldsymbol{\tau}^2_{\beta_1}\end{bmatrix}, \qquad \boldsymbol{M}_{j|0}=\begin{bmatrix}\mu_\alpha\\ z_j\boldsymbol{\mu}_{\beta_1}\end{bmatrix},
$$

$$
\boldsymbol{T}^*_{j|0}=\left(\boldsymbol{X}'_{j|0}\Omega_j^{-1}\boldsymbol{X}_{j|0}+\boldsymbol{T}^{-1}_{j|0}\right)^{-1} \quad \text{and} \quad \boldsymbol{M}^*_{j|0}=\boldsymbol{T}^*_{j|0}\left(\boldsymbol{X}'_{1|0}\Omega_j^{-1}\boldsymbol{y}_j+\boldsymbol{T}^{-1}_{j|0}\boldsymbol{M}_{j|0}\right)
$$

(b) and

$$
\begin{aligned}
p(\boldsymbol{y}_j|\Omega_j,\Theta_1,c_j=t,S_j=1) &= \frac{|\boldsymbol{T}^*_{j|1}|^{1/2}|}{(2\pi)^{T/2}|\boldsymbol{T}_{j|1}|^{1/2}|\Omega_j|^{1/2}} \\
&\times\exp\left\{-\frac{1}{2}\left(\boldsymbol{y}_j'\Omega_j^{-1}\boldsymbol{y}_j+\boldsymbol{M}'_{j|1}\boldsymbol{T}^{-1}_{j|1}\boldsymbol{M}_{j|1}-\boldsymbol{M}^{*\prime}_{j|1}\boldsymbol{T}^{*-1}_{j|1}\boldsymbol{M}^*_{j|1}\right)\right\} \\
&\times\frac{\Phi\left((\infty,\infty,0)'|\boldsymbol{M}^*_{1|j},\boldsymbol{T}^*_{1|j}\right)-\Phi\left((\infty,0,0)'|\boldsymbol{M}^*_{1|j},\boldsymbol{T}^*_{1|j}\right)}{\Phi\left((\infty,\infty,0)'|\boldsymbol{M}_{1|j},\boldsymbol{T}_{1|j}\right)-\Phi\left((\infty,0,0)'|\boldsymbol{M}_{1|j},\boldsymbol{T}_{1|j}\right)}
\end{aligned}
$$

where

$$\boldsymbol{T}_{j|1} = \begin{bmatrix} \tau_\alpha^2 & 0 & 0 \\ 0 & z_j\tau_{\beta_1}^2 & 0 \\ 0 & 0 & z_j\tau_{\beta_2}^2 \end{bmatrix}, \qquad \boldsymbol{M}_{j|1} = \begin{bmatrix} \mu_\alpha \\ z_j\mu_{\beta_1} \\ z_j\mu_{\beta_2} \end{bmatrix},$$

$$\boldsymbol{T}^*_{j|1} = \left(\boldsymbol{X}'_{j|1}\Omega_j^{-1}\boldsymbol{X}_{j|1} + \boldsymbol{T}^{-1}_{j|1}\right)^{-1} \quad \text{and} \quad \boldsymbol{M}^*_{j|1} = \boldsymbol{T}^*_{j|1}\left(\boldsymbol{X}'_{j|1}\Omega_j^{-1}\boldsymbol{y}_j + \boldsymbol{T}^{-1}_{j|1}\boldsymbol{M}_{j|1}\right)$$

2. Sample $\boldsymbol{C}$.

   Draw $\boldsymbol{C}$ from

$$p(\boldsymbol{C}|\boldsymbol{Y},\boldsymbol{\Theta},\boldsymbol{S},\boldsymbol{\Omega}) = \prod_{j=1}^{J} p(c_j = t|\boldsymbol{\Theta},\boldsymbol{y}_j,S_j,\Omega_j)$$

   If $S_j = 0$, c_j no sampling is required. Conditional on $S_j = 1$, c_j is drawn according to

$$p(c_j = t|\boldsymbol{\Theta},\boldsymbol{y}_j,S_j = 1) \quad = \quad \frac{\frac{1}{T-2}p(\boldsymbol{y}_j|\Theta_1,c_j = t,S_j = 1,\Omega_j)}{\sum_{t'=1}^{T-2}\frac{1}{T-2}p(\boldsymbol{y}_j|\Theta_1,c_j = t',S_j = 1,\Omega_j)}$$

   where the densities in the denominator and numerator are given in step 1.

3. Sample $\boldsymbol{B}$.

   Draw $\boldsymbol{B}$ from

$$p(\boldsymbol{B}|\boldsymbol{Y},\boldsymbol{\Theta},\boldsymbol{S},\boldsymbol{C},\boldsymbol{\Omega}) = \prod_{j:S_j=0} p(\boldsymbol{b}_{j|0}|\boldsymbol{y}_j,\Theta_0,S_j = 0,\Omega_j) \prod_{j:S_j=1} p(\boldsymbol{b}_{j|1}|\boldsymbol{y}_j,\Theta_1,S_j = 1,c_j = t,\Omega_j)$$

   Again, from step 1 we can see that $\boldsymbol{b}_{j|0}$ is drawn according to $N(\boldsymbol{M}^*_{j|0},\boldsymbol{T}^*_{j|0})$ restricted to the region $\mathbb{R}\times C_+$ and $\boldsymbol{b}_{j|1}$ is sampled according to $N(\boldsymbol{M}^*_{j|1},\boldsymbol{T}^*_{j|1})$ restricted to the region $\mathbb{R}\times C_+\times C_-$. To draw $\boldsymbol{b}_{j|0}$ and $\boldsymbol{b}_{j|1}$ we note that linear transformations of truncated normal vectors, and the one-dimensional conditional distributions, are also truncated normal (Rodriguez-Yam, Davis, & Scharf, 2004), so that drawing the elements of $\boldsymbol{b}_{j|0}$ and $\boldsymbol{b}_{j|1}$, reduces to drawing a sequence of one-dimensional constrained conditional normal distributions.

4. Sample $\boldsymbol{\lambda}$.

If the basis functions are exponential growth curves then draw $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_J)$, from

$$\begin{aligned} p(\boldsymbol{\lambda}|\boldsymbol{Y},\boldsymbol{S},\boldsymbol{C},\boldsymbol{B},\boldsymbol{\Omega},\mu_\lambda,\tau^2_\lambda) &= \prod_{j=1}^{J} p(\lambda_j|\boldsymbol{y}_j, S_j = s_j, c_j, \boldsymbol{b}_{j|s_j}, \Omega_j, \mu_\lambda, \tau^2_\lambda) \\ &= \prod_{j=1}^{J} p(\boldsymbol{y}_j|\lambda_j, S_j = s_j, c_j, \boldsymbol{b}_{j|s_j}, \Omega_j) p(\lambda_j|\mu_\lambda, \tau^2_\lambda) \end{aligned}$$

using a Metropolis-Hastings step. If the current value of λ_j in the chain is λ_j^c then a new value, λ_j^N, is drawn from a proposal density $q(\lambda_j) \sim N_{C_\lambda}(\hat{\lambda}_j, \hat{\Sigma}_{\lambda_j})$. The value of $\hat{\lambda}_j$ is the value that maximizes $l(\lambda_j)$ where $l(\lambda_j) = \log(p(\boldsymbol{y}_j|\lambda_j, S_j = s_j, c_j, \boldsymbol{b}_{j|s_j}, \Omega_j)p(\lambda_j|\mu_\lambda, \tau^2_\lambda))$, and $\hat{\Sigma}_{\lambda_j}$ is equal to the inverse of the second derivative of $l(\lambda_j)$ evaluated at $\hat{\lambda}_j$. If $\lambda_j^N > 0$, λ_j^N is accepted with the usual Metropolis-Hastings probability, otherwise retain λ_j^c.

5. Sample (σ^2_E, σ^2_I).

   (a) When $\varepsilon_{tj} \sim N(0, (\sigma^2_E, \sigma^2_I)\boldsymbol{z}_j)$ then draw (σ^2_E, σ^2_I) from

$$p(\sigma^2_E, \sigma^2_I|\boldsymbol{Y},\boldsymbol{B},\boldsymbol{S},\boldsymbol{C}) = p(\sigma^2_E|\boldsymbol{Y},\boldsymbol{B},\boldsymbol{S},\boldsymbol{C})p(\sigma^2_I|\boldsymbol{Y},\boldsymbol{B},\boldsymbol{S},\boldsymbol{C})$$

   where

$$\sigma^2_E \sim IG\left(\frac{J_E}{2} - 1, \frac{\sum_{\{j:\boldsymbol{z}_j=(1,0)'\}}(\boldsymbol{y}_j - \hat{\boldsymbol{y}}_j)'(\boldsymbol{y}_j - \hat{\boldsymbol{y}}_j)}{2}\right)\mathbb{I}\{\sigma^2_E \le k\},$$

$$\mathbb{I}\{\sigma^2_E \le k\} = \begin{cases} 0 & \text{if} \sigma^2_E > k \\ 1 & \text{if} \sigma^2_E \le k, \end{cases}$$

$$\hat{\boldsymbol{y}}_j = \begin{cases} \boldsymbol{X}_{j|0}\boldsymbol{b}_{j|0} & \text{if} \quad S_j = 0 \\ \boldsymbol{X}_{j|1}\boldsymbol{b}_{j|1} & \text{if} \quad S_j = 1 \quad \text{and} \quad c_j = t^* \end{cases}$$

   and $J_E = \sum_{j=1}^{J} \mathbb{I}\{\boldsymbol{z}_j = (1,0)'\}$. Similarly, draw σ^2_I with $\boldsymbol{z}_j = (0,1)'$.

   (b) If $\varepsilon_{jt} \sim \sigma_j t_3$ then draw σ^2_j by

      i. Generating κ_{tj}, from a Gamma distribution $G(u_a, u_b)$ with $u_a = 2$ and

$$u_b = \frac{1}{2}\left(1 + \left(\frac{y_{tj} - \boldsymbol{X}_{tj|S_j}\boldsymbol{b}_{j|S_j}}{\sigma\sqrt{3}}\right)^2\right)$$

      where $\boldsymbol{X}_{tj|S_j}$ is a row vector denoting the tth row of $\boldsymbol{X}_{j|S_j}$ for $t = 1, \ldots, T$ and $j = 1, \ldots, J$.

ii. Generating $\sigma^2 = (\sigma_E^2, \sigma_I^2) z_j$. σ_E^2 and σ_I^2 have inverse gamma distribution with parameters (u_E, v_E) and (u_I, v_I) respectively. To draw σ_E^2, we note $u_E = J_E/2 - 1$ where $J_E = \sum_{j=1}^{J} \mathbb{I}\{z_j = (1,0)'\}$ and

$$v_E = \frac{1}{2} \sum_{\{j:\mathbf{z}_j=(1,0)'\}} \sum_{t=1}^{T} \left(\frac{y_{tj} - \boldsymbol{X}_{tjS_j} \boldsymbol{b}_{jS_j}}{\sqrt{\kappa_{tj}/3}} \right)^2$$

σ_I^2 is drawn in a similar fashion.

6. Sample $\boldsymbol{\delta} = (\delta_0, \delta_1)$.

Draw $\boldsymbol{\delta}$ from

$$p(\boldsymbol{\delta}|\boldsymbol{Y}, C, \boldsymbol{B}, \boldsymbol{S}) = p(\boldsymbol{\delta}|\boldsymbol{S}) \propto p(\boldsymbol{S}|\boldsymbol{\delta}) p(\boldsymbol{\delta}),$$

where $p(\boldsymbol{\delta})$ is the prior distribution of $\boldsymbol{\delta}$ discussed in the main text. We use a Metropolis-Hastings method for this step . If the current value of $\boldsymbol{\delta}$ in the chain is $\boldsymbol{\delta}^c$ then a new value, $\boldsymbol{\delta}^N$, is drawn from a proposal density $q(\boldsymbol{\delta}) \sim N(\hat{\boldsymbol{\delta}}, \hat{\Sigma})$, where $\hat{\boldsymbol{\delta}}$ is the value of $\boldsymbol{\delta}$ which maximizes $\log[p(\boldsymbol{S}|\boldsymbol{\delta})p(\boldsymbol{\delta})]$, and $\hat{\Sigma}$ is equal to the inverse of the second derivative of $\log[p(\boldsymbol{S}|\boldsymbol{\delta})p(\boldsymbol{\delta})]$ evaluated at $\hat{\boldsymbol{\delta}}$. This new value is accepted with the usual probability.

7. Sample $(\mu_\alpha, \boldsymbol{\mu}_{\beta_1}, \boldsymbol{\mu}_{\beta 2})$.

First, draw μ_α from

$$\mu_\alpha | \boldsymbol{B}, \tau_\alpha^2 \sim N\left(\frac{\tau_\alpha^2 g_\alpha + h_\alpha \sum_{j=1}^{J} \alpha_j}{J \times h_\alpha + \tau_\alpha^2}, \frac{J \times h_\alpha + \tau_\alpha^2}{\tau_\alpha^2 h_\alpha} \right)$$

then draw $(\mu_{\beta_{1E}}, \mu_{\beta_{2E}})$ from

$$p(\mu_{\beta_{1E}}, \mu_{\beta_{2E}} | \boldsymbol{Y}, \boldsymbol{B}, \tau_{\beta_{1E}}^2, \tau_{\beta_{2E}}^2) = p(\mu_{\beta_{1E}} | \boldsymbol{B}, \tau_{\beta_{1E}}^2) \times p(\mu_{\beta_{2E}} | \boldsymbol{B}, \tau_{\beta_{2E}}^2)$$

where

$$\mu_{\beta_{1E}}|\boldsymbol{B},\boldsymbol{\tau}^2_{\beta_1 E} \sim N_{C_+}\left(\frac{\tau^2_{\beta_{1E}} g_{\beta_1} + h_{\beta_1}\sum_{\{j:\boldsymbol{z}_j=(1,0)\}}\beta_{1j}}{J_E h_{\beta_1}+\tau^2_{\beta_{1E}}}, \frac{J_E h_{\beta_1}+\tau^2_{\beta_{1E}}}{\tau^2_{\beta_{1E}} h_{\beta_1}}\right)$$

$$\mu_{\beta_{2E}}|\boldsymbol{B},\boldsymbol{\tau}^2_{2E} \sim N_{C_-}\left(\frac{\tau^2_{\beta_{2E}} g_{\beta_2} + h_{\beta_2}\sum_{\{j:\boldsymbol{z}_j=(1,0),S_j=1\}}\beta_{2j}}{J_{E_s} h_{\beta_2}+\tau^2_{\beta_{2E}}}, \frac{J_{E_s} h_{\beta_2}+\tau^2_{\beta_{2E}}}{\tau^2_{\beta_{2E}} h_{\beta_2}}\right),$$

$J_E = \sum_{j=1}^J \mathbb{I}\{\boldsymbol{z}_j = (1,0)'\}$ and $J_{E_s} = \sum_{j=1}^J \mathbb{I}\{\boldsymbol{z}_j = (1,0)', S_j = 1\}$. Then draw $(\mu_{0I}, \mu_{\beta_{1I}}\mu_{\beta_{2I}})'$ in a similar fashion but $\boldsymbol{z}_j = (0,1)$.

8. Sample $(\tau^2_\alpha, \tau^2_{\beta_{1I}}, \tau^2_{\beta_{2I}}, \tau^2_{\beta_{1E}}, \tau^2_{\beta_{2E}})'$.

   First, draw τ^2_α from $p(\tau^2_\alpha|\boldsymbol{B},\mu_\alpha)$, then draw, $(\tau^2_{\beta_{1E}}, \tau^2_{\beta_{2E}})'$ from

$$p(\tau^2_{\beta_{1E}}, \tau^2_{\beta_{2E}}|\boldsymbol{Y},\boldsymbol{B},\mu_{\beta_{1E}},\mu_{\beta_{2E}}) = p(\tau^2_{\beta_{1E}}|\boldsymbol{B},\mu_{\beta_{1E}}) \times p(\tau^2_{\beta_{2E}}|\boldsymbol{B},\boldsymbol{\mu}_{\beta_{2E}}).$$

   where

$$\tau^2_\alpha|\boldsymbol{B},\mu_\alpha \sim IG\left(J/2-1, \frac{\sum_{j=1}^J(\alpha_j-\mu_\alpha)^2}{2}\right)\mathbb{I}\{\tau^2_\alpha \le a_\alpha\}$$

$$\tau^2_{1E}|\boldsymbol{B},\mu_{\beta_{1E}} \sim IG\left(J_E/2-1, \frac{\sum_{\{j:\boldsymbol{z}_j=(1,0)\}}(\beta_{1j}-\mu_{\beta_{1E}})^2}{2}\right)\mathbb{I}\{\tau^2_{1E} \le a_{\beta_1}\}$$

$$\tau^2_{2E}|\boldsymbol{B},\mu_{\beta_{2E}} \sim IG\left(J_{E_s}/2-1, \frac{\sum_{\{j:\boldsymbol{z}_j=(1,0),S_j=1\}}(\beta_{2j}-\mu_{\beta_{2E}})^2}{2}\right)\mathbb{I}\{\tau^2_{1E} \le a_{\beta_1}\}$$

   where J_E, J_{E_s} are as defined in step 7, the function $\mathbb{I}\{\cdot\}$ is as defined in step 5 and $a_\alpha, a_{\beta_1}, a_{\beta_2}$ are calculated as described in the Priors section. Then draw $(\tau^2_{\beta_{1I}}, \tau^2_{\beta_{2I}})$ in a similar fashion but with $\boldsymbol{z}_j = (0,1)$.

9. If the basis functions are exponential growth curves then μ_λ and τ^2_λ are drawn as in steps 7 and 8 above with the appropriate constraints.

TABLE 3

Estimate of posteriors means for individual probability of spiralling, $\hat{\Pr}(S_j = 1|\boldsymbol{Y})$ for all individuals classified as entity theorists (red) and as incremental theorists (blue) for three basis functions and two type of error distribution. An * or * indicates an individual classified as an entity theorist or incremental theorist respectively for whom the probability of spiralling is greater than 0.5. An estimate of the median value of the point at which the spiral begins, c_j, is given in the last column for the case when $f_j(t) = 1 - \exp\{-\lambda_j t\}$.

	$f(t) = t$		$f_j(t) = 1 - \exp\{-\lambda_j t\}$		$\hat{c}_j$
Individual	Normal	t_3	Normal	t_3	
1	0.06	0.06	0.01	0.01	0
2	0.88 *	0.44	0.99 *	0.97 *	5
3	0.27	0.28	0.22	0.09	0
4	0.03	0.06	0.01	0.01	0
5	0.16	0.25	0.20	0.09	0
6	0.01	0.03	0.00	0.00	0
7	0.05	0.08	0.01	0.01	0
8	0.10	0.11	0.02	0.03	0
9	0.01	0.04	0.01	0.00	0
10	0.20	0.37	0.22	0.07	0
11	0.06	0.04	0.00	0.00	0
12	0.12	0.13	0.02	0.03	0
13	0.11	0.24	0.15	0.04	0
14	0.76 *	0.68 *	0.77 *	0.93 *	4
15	0.22	0.22	0.05	0.14	0
16	0.55 *	0.58 *	0.54 *	0.66 *	4
17	0.05	0.12	0.02	0.02	0
18	0.98*	0.87*	0.97*	1.00*	3
19	0.97*	0.46*	0.66*	0.07	0
20	0.96*	0.86*	0.95*	0.99*	3
21	1.00*	0.92*	1.00*	0.97*	4
22	0.99*	0.97*	1.00*	1.00*	3
23	0.01	0.05	0.01	0.01	0
24	0.59*	0.66*	0.59*	0.75*	3
25	0.03	0.11	0.01	0.02	0
26	1.00*	0.98*	1.00*	1.00*	2
27	0.04	0.23	0.01	0.01	0
28	0.97 *	0.97 *	1.00 *	0.94 *	1
Average	0.63	0.68	0.64	0.61	
Average	0.17	0.19	0.14	0.14	

TABLE 4
Values of π_E and π_I used in simulation settings.

Parameter	Setting Number		
	1	2	3
π_E	0.0	0.5	0.6
π_I	0.0	0.5	0.1
$\pi_E - \pi_I$	0.0	0.0	0.5

FIGURE 7

Estimated posterior densities for the model $f_j(t) = 1 - \exp\{-\lambda_j t\}$ and $\varepsilon \sim \sigma_j t_3$. Panel (a) displays the difference in the probability of spiralling between entity theorists and incremental theorists, $\pi_E - \pi_I$. Panel (b) shows the difference in maximal performance gain between entity and incremental theorists, $\mu_{\beta_{1E}} - \mu_{\beta_{1I}}$.

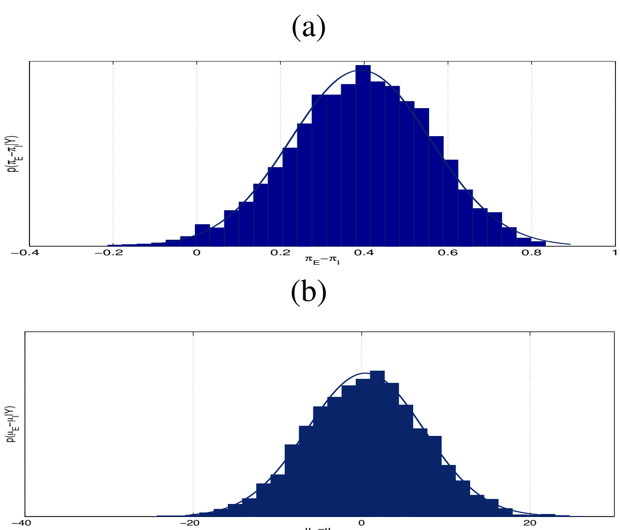


FIGURE 8

Panel (a); Posterior mean of all individual performance curves for entity theorists (red) and incremental theorists (blue) for the model with $f_j(t) = 1 - \exp\{-\lambda_j t\}$ and $\varepsilon_{jt} \sim \sigma t_3$. Panels (b) and (c) are similar plots for individuals for whom the probability of spiralling is less than 0.5 (panel (b)) and greater than 0.5 (panel (c)).

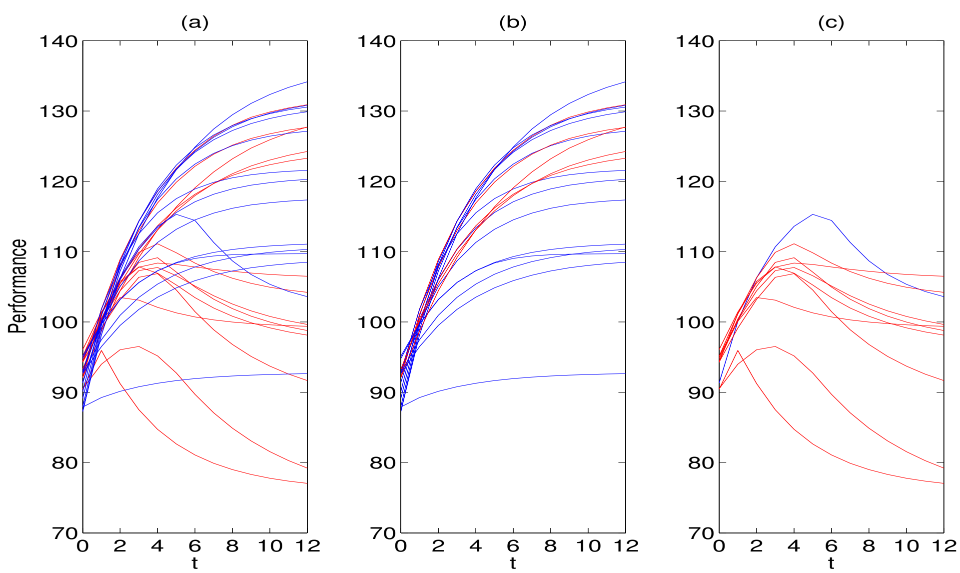


FIGURE 9

Observed performance of individual 19 and posterior mean of regression line when $\varepsilon_{jt} \sim N(0, \sigma_j^2)$, dashed (- - -), and when $\varepsilon_{tj} \sim \sigma_j t_3$, dotted (...), for $f(t)_j = 1 - \exp\{-\lambda_j t\}$.

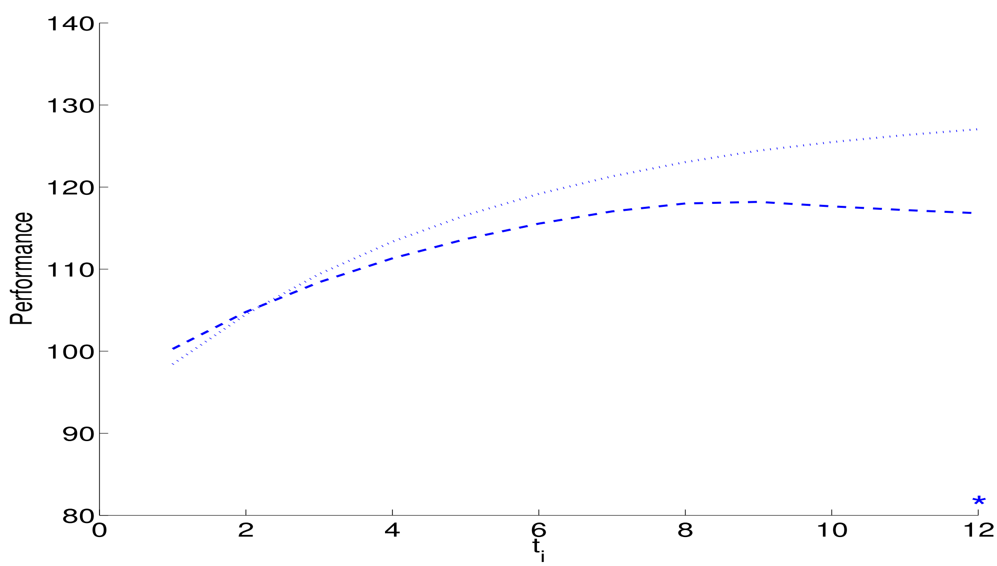


FIGURE 10

Histogram of residuals for the model given by (.4) and (.5) with $\varepsilon_{jt} \sim \sigma_j \times t_3$, and $f(t)_j = 1 - \exp\{-\lambda_j t\}$, overlaid with the density function of a t_3.

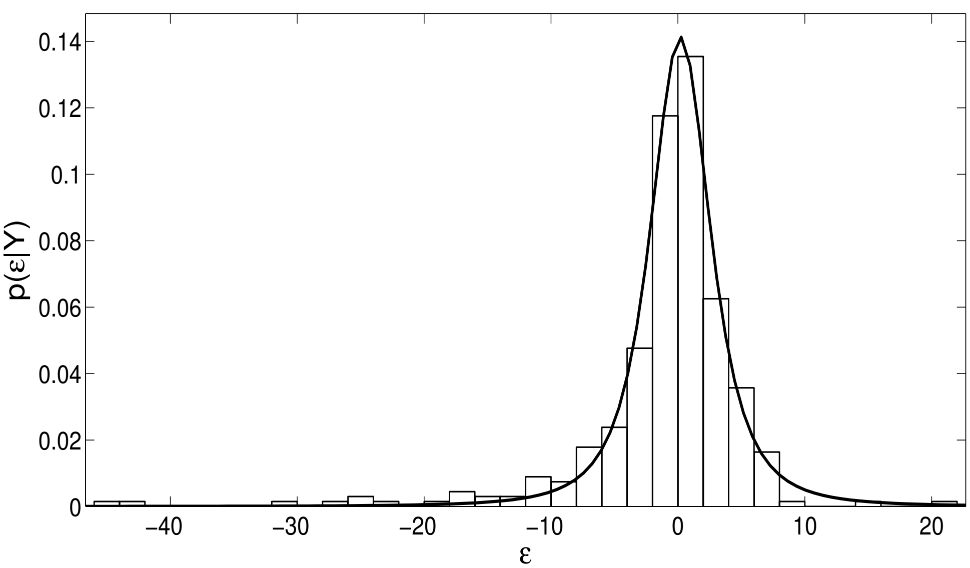


FIGURE 11

Boxplots of posterior mean estimates for 3 simulation settings with 50 realisations in each simulation. In each panel, the left boxplot corresponds to the simulation when $\pi_E = \pi_I = 0$, the middle boxplot corresponds to the simulation when $\pi_E = \pi_I = 0.5$ and the right boxplot corresponds to the simulation when $\pi_E = 0.6$ and $\pi_I = 0.1$. Panel (a) reports posterior mean estimates of $\pi_E - \pi_I$. Panel (b) reports posterior mean estimates of $\mu_{\beta_{1E}} - \mu_{\beta_{1I}}$, Panel (c) reports posterior mean estimates of σ_E/σ_I. The horizontal blue dashed line is true values.

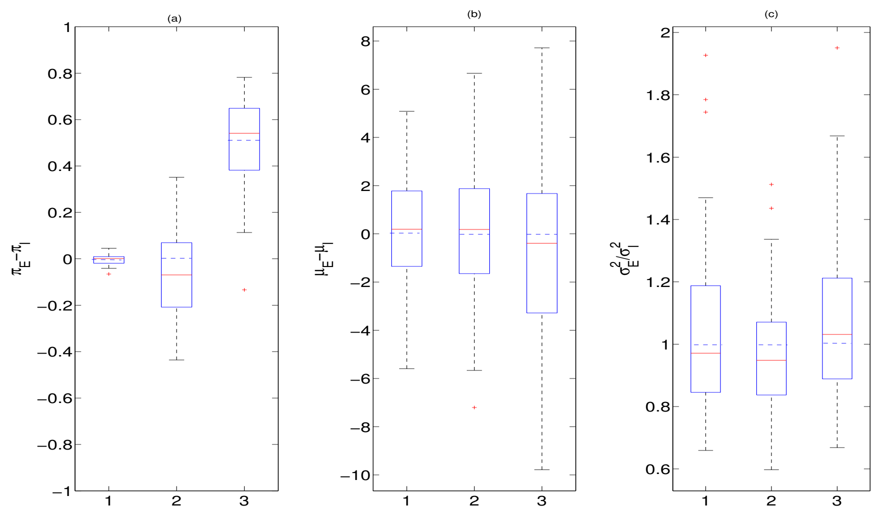


FIGURE 12

Effect of c_δ on the prior for $\boldsymbol{\pi} = (\pi_E, \pi_I)$. In panel (a), $c_\delta = 1$, in panel (b) $c_\delta = 4$, and in panel(c) $c_\delta = 10$.