 doi:10.20944/preprints202207.0246.v1

Article

Motivated, Willing, and Able: Non-Cognitive Factors Influence Complex Problem Solving Performance beyond Cognitive Ability

Abstract: Complex problem-solving (CPS) tasks have become an increasingly popular tool for understanding and assessing cognitive ability. These tasks have been repeatedly shown to be predictors of academic and workplace success above and beyond traditional measures of general intelligence and fluid intelligence. To date, there has been little exploration of the underlying mechanisms that drive this additional predictive utility. In this study, we examined the role of a variety of non-cognitive personality and investment traits that could drive performance on CPS tasks. Adult participants (n = 152) were recruited via M-Turk and completed a battery of personality and investment trait measures, a measure of general mental ability, and a 61-trial microworlds-style CPS task. Generalised linear mixed-effects models revealed a wide variety of personality and investment traits influenced task performance above and beyond general mental ability. Specifically, two clusters of traits emerged as important determinants of performance: one cluster that influenced the capacity to deal with the introduction of system randomness (Conscientiousness and Extraversion) and one cluster that influenced the capacity to deal with the introduction of system delays (NFC, Learning Goal Orientation, and Intellect). These findings suggest that CPS tasks do capture more than just general mental ability and may be good predictors of academic and workplace success because they tap into both cognitive ability and the motivation and willingness to engage in cognitive exploration and mental effort.

Keywords: complex problem solving; microworlds; personality; investment traits; within-individual variability; performance trajectories

1. Introduction

In the last two decades, complex problem-solving (CPS) tasks have become an increasingly popular form of cognitive assessment (Lotz et al., 2016; Wüstenberg et al., 2012). CPS tasks are purported to represent the capacity to analyse and solve novel problems, a critical skill for educational and occupational success (Lotz et al., 2016; Stadler et al., 2019). In CPS tasks, the respondent must learn to control a system with at least one input and at least one output over an extended series of trials (Dörner & Funke, 2017; Funke, 2010). These systems are often embedded in realistic contexts, for example, inventory management for business professionals (Birney et al., 2018), playing handball for children and adolescents (Lotz et al., 2016), or in more constrained contexts such as 3-input/3-output deterministic systems (Beckmann et al., 2017; Burns & Vollmeyer, 2002). Knowledge of the relationships between inputs and outputs must be elicited through exploration, then the participant must apply this knowledge to control the system (Dörner & Funke, 2017; Funke, 2010). CPS tasks were thus, in addition to understanding problem-solving processes, designed to reflect the extended, dynamic, and technological problem-solving challenges that people encounter at school, university, and the workplace (Beckmann et al., 2017; Funke, 2001; Rigas & Brehmer, 1999).

A significant point of contention within the field of CPS research is whether CPS tasks capture cognitive or non-cognitive processes beyond those captured by tests of general intelligence (g) or the related broad ability, fluid intelligence (Gf) (Greiff & Neubert, 2014; Lotz et al., 2016; Stadler et al., 2015; Wüstenberg et al., 2012). Theoretically, the design and demands of CPS tasks are intended to reflect the kind of intelligence required to succeed in everyday life (Dörner & Funke, 2017; Funke, 2010), as opposed to common tests of

 doi:10.20944/preprints202207.0246.v1

g or Gf used in assessment and selection, which are often conceptually and visually different to the situations people encounter day-to-day (Dörner & Funke, 2017; Funke, 2010). The majority of research to date supports this assertion, as CPS performance explains variation beyond both g and Gf in academic performance (Greiff et al., 2015; Greiff et al., 2012; Kretzschmar et al., 2016; Lotz et al., 2016; Sonnleitner et al., 2013; Wüstenberg et al., 2012) and supervisor-rated workplace performance (Danner et al., 2011).

If CPS tasks explain more variation in practical outcomes than tests of general or broad cognitive ability, this implies that CPS tasks may capture multiple cognitive and/or non-cognitive processes. But what are these processes? Answering this question is vital for the accurate interpretation and use of results from CPS tasks. Previous studies have largely examined how cognitive and non-cognitive processes affect overall performance; few studies have examined how these factors dynamically influence performance throughout the task (an exception being Birney et al., 2018, who found mixed evidence for the role of non-cognitive factors in CPS performance). Therefore, the purpose of this study is to begin developing a process-driven account of CPS task performance. Specifically, we focus on how non-cognitive processes are invoked during a challenging and confusing CPS task. We leverage generalised linear mixed-effects modelling techniques to understand how these processes differentially influence performance as the task becomes more challenging and as people gain more experience.

CPS Task Demands on Cognitive and Non-Cognitive Capacities

Early CPS tasks were complex, dynamic, and highly realistic microworld simulations that often contained hundreds of variables to be controlled (Dörner & Funke, 2017; Funke, 2010). Dörner (1986, as cited in Funke, 2010) conceived the term 'operative intelligence' to reflect CPS task performance being multiply determined by "… strategic processes like flexibility, foresight, circumspection, or systematic behaviour" (Funke, 2010, p. 135). Modern microworlds simulations remain complex, dynamic, and realistic, but often contain fewer variables (Birney et al., 2018; Gonzalez et al., 2005; Wood et al., 2009). Dörner (1986, as cited in Funke, 2010) proposed that microworld simulations demanded four concurrent processes: 1) acquiring knowledge of the system and developing a mental model that integrates this knowledge, 2) decomposing goals into attainable steps and balancing competing goals, 3) forward planning and decision making, and 4) self-management of task-induced affect. We note two interesting aspects of this account that are particularly relevant to the current study. Firstly, CPS tasks require two phases of cognitive control: learning how the system works (process 1), then applying this learning to achieve goals (processes 2-4). Secondly, the complexity and duration of CPS tasks invokes substantial metacognitive demands to manage competing goals, plan future steps, and self-monitor and regulate (processes 2-4). Successful control of a CPS task thus also requires the engagement of non-cognitive processes.

Performance on challenging cognitive tasks has different relations with intelligence and personality depending on the situation in which the task is completed; such intelligence-personality-performance relations are outlined in detail by von Stumm et al. (2011). Strong situations encompass maximal performance environments in which people exert significant cognitive effort to perform their best. In strong situations, cognitive test performance is strongly related to intelligence but weakly related to personality. In contrast, in weak situations cognitive test performance is strongly related to personality but weakly related to intelligence. When the stakes are lower and the task is longer, such as performing typical workplace duties, successful task performance is most heavily influenced by personality and motivational dispositions that dictate the amount of discretionary cognitive effort invested in the task.

CPS tasks tend to require extended time periods of cognitive effort – like the kind of tasks typically demanded in education and the workplace (Beckmann et al., 2017; Funke, 2001; Rigas & Brehmer, 1999) – and thus are more comparable to weak situations than strong situations. Consequently, we could expect CPS task performance to be driven by both cognitive and non-cognitive processes, such as personality, metacognition, or motivation. Furthermore, the relationship between cognitive and non-cognitive factors, and CPS task performance may change in strength and direction at different phases of the task, as people gain more experience with the system within phases, but also as the task becomes more or less challenging from one phase to the next. As significant cognitive effort cannot be sustained for an extended period of time by many people, non-cognitive factors may become the stronger influence on performance in later phases particularly if they are more challenging. This is also consistent with the early work of Ackerman (1987, 1988), who demonstrated changes in the importance of different individual differences factors as task knowledge is acquired and expertise develops.

Given that CPS tasks explain variation in real world outcomes beyond traditional intelligence measures, successful performance on CPS tasks may be sensitive to non-cognitive processes—in addition to general mental ability—that drive success on temporally extended and dynamic cognitive challenges. We investigate two related yet differentiable groups of non-cognitive traits that could be responsible for this additional variation in practical outcomes: intellectual investment traits and personality traits.

Intellectual Investment Traits

Intellectual investment traits are stable individual differences in one's propensity to engage in cognitively challenging activities (von Stumm & Ackerman, 2013; von Stumm et al., 2011). Ackerman (1996) proposed that these traits causally influence the development of adult cognitive abilities by facilitating the transition from intelligence-as-process to intelligence-as-knowledge via two pathways. First, intellectual investment traits motivate individuals to proactively seek out opportunities for cognitive challenge throughout their development. Second, these traits predispose individuals to invest more effort in cognitive challenges when they are encountered.

We may expect these traits to influence CPS performance in a number of ways. For instance, by virtue of being motivated to seek cognitive challenges, people high in these traits will have more experience and skills in problem-solving, which will confer baseline performance improvements. They are also more likely to be driven to engage with and master CPS tasks as an intellectual activity in and of itself, and thus perform better as a result. People high in intellectual investment traits may also be more adept at self- and task-monitoring, which would facilitate more efficient and effective acquisition of task rules and control performance, particularly in high complexity phases of the task.

The literature reveals small relationships between CPS performance and several intellectual investment traits. These include Goal Orientation (Birney et al., 2018; Cripps et al., 2016; Vollmeyer & Rheinberg, 1999), Implicit Theories of Ability (Birney et al., 2018; Cripps et al., 2016), Cognitive Reflection (Hundertmark et al., 2015), and Need for Cognition (NFC) (Rudolph et al., 2018). Only two of these studies found that the traits significantly influenced CPS performance when controlling for g or Gf. Performance Goal Orientation had a positive effect on performance in less complex CPS task variants in Birney et al. (2018). In Hundertmark et al. (2015), Cognitive Reflection was a stronger predictor of overall CPS performance than g but this effect was reduced in more complex variants of the task.

In the current study, we examine the influence of three intellectual investment traits: NFC (Cacioppo & Petty, 1982), Goal Orientation (VandeWalle, 1997), and Intellect (Mussel, 2013). Need for Cognition is a prototypical investment trait, representing an individual's motivation to seek out and engage in cognitively challenging activities (Cacioppo & Petty, 1982). Goal Orientation represents an individual's preference for striving for various types of goals (VandeWalle, 1997). It is segmented into Learning Goal Orientation, the preference for goals that develop deep learning and mastery of skills and knowledge, and Performance Goal Orientation, the preference for goals that lead to (or allow for) performance comparisons against some standard (VandeWalle, 1997). Performance Goal Orientation is further segmented into Performance-Prove and Performance-Avoid orientations to reflect the motivation for striving towards a standard, to prove themselves by being favourably appraised, versus avoiding negative appraisals, respectively (VandeWalle, 1997). Intellect, as conceptualised by Mussel (2013), is theoretically similar to NFC in that it captures an individual's willingness to engage in challenging mental activities but provides theoretically interesting measurement features. In Mussel's (2013) theory, Intellect is segmented into two processes and three operations. The Intellect processes are Seek (seeking out opportunities for intellectual engagement) and Conquer (conquering intellectual opportunities during the task). The Intellect operations reflect three different domains of intellectual engagement: Thinking (related to Gf), Learning (related to crystallised intelligence (Gc)), and Creating (related to creativity). Intellect is thus a useful non-cognitive construct that may be related to important processes that are relevant to CPS.

Personality

While some personality traits are considered intellectual investment traits, particularly the Intellect aspect of Openness/Intellect, we could theoretically expect CPS performance to be related to other traits as well. Nevertheless, evidence for the relationship between personality and CPS performance is sparse. In a study by Greiff and Neubert (2014) in an adolescent sample, correlations between CPS performance and personality were small but significant; there were small negative relationships between Conscientiousness (β = -.10 for both knowledge acquisition and knowledge application performance), Neuroticism (β = -.12 for knowledge acquisition and β = -.14 knowledge application), Agreeableness (β = -.21 for knowledge acquisition and β = -.18 knowledge application). There were no relationships with Extraversion or Openness to Experience, although the authors noted the poor internal reliability of the scale in their sample. In another study on adult management professionals attending a training course, there were no significant relationships with any measured personality trait (Birney et al., 2018).

Although this suggests relations between personality and performance may be small, we still consider it pertinent to re-investigate these associations. Both of the aforementioned studies examined higher trait-level personality associations only. In the current study, however, we leverage a facet-level measure of the Five Factor Model, the Big Five Inventory-2 (BFI-2), which segments each personality trait into three facets (Soto & John, 2017). Within each trait, facets can be differentially related to different outcomes and behaviours (Soto & John, 2017), thus domain-level analysis can provide more nuanced understanding of personality-performance relations. Furthermore, in the current study we examine a broader adult sample drawn from the community, rather than a restricted range as in previous studies, thus we expect more variation in self-reported personality and investment traits.

The present study

The present study aims to develop a process account of CPS performance. We reanalyse a data set that was partially analysed in Study 5 of Fayn et al. (2019). We focus on the influence of personality and intellectual investment traits on a CPS microworlds task designed to be challenging, confusing, and complex. Specifically, we are interested in how these traits differentially affect performance trajectories over time. This analysis is enabled by generalised linear mixed-effects modelling, which allows us to identify and compare performance trajectories on our task as a function of cognitive and non-cognitive traits, and task complexity levels. This study makes three key novel contributions.

The first novel contribution of this study is the use of generalised linear mixed-effects modelling to understand how cognitive and non-cognitive factors influence both overall task performance and item-to-item performance trajectories (Birney et al., 2019). CPS tasks require extended engagement and capture one's capacity to both learn and control a complex and dynamic problem (Dörner & Funke, 2017; Funke, 2010). This means that traditional statistical analysis methods using aggregate or mean scores are limited in their capacity to provide an explanatory account of CPS performance (Birney et al., 2019). CPS tasks are better suited to item-to-item statistical analysis methods that allow us to understand performance trajectories throughout the task. This type of analysis enables us to develop a nuanced process-driven account of how individuals monitor and control a CPS

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

microworld, and how non-cognitive factors influence differences in performance trajectories. We examine this relationship across the three stages of the task, each of increasing complexity.

Second, we extend upon previous findings by examining how lower-level personality trait facets dynamically influence performance in a broad adult sample. While one previous study has applied generalised linear mixed-effects modelling to examine the influence of cognitive and non-cognitive factors on CPS performance, we note that this study used a range-restricted sample of adults in management roles and examined facet-level personality traits (Birney et al., 2018). We use a personality measure that allows us to examine facet-levels, which is critical as different facets can have quite different relationships with cognitive performance and cognitive challenge-seeking behaviours (Soto & John, 2017).

Finally, the CPS microworld task used has been modified to include a dynamic within-block manipulation of complexity. Previous studies (Birney et al., 2018) have manipulated complexity across blocks. This addition of dynamic change in complexity while problem-solving adds demand for a sensitivity to feedback (i.e., a change in an input-output contingency), as well as tolerance to potential confusion that this generates.

Aims and Hypotheses

In this study, we aim to better understand the way people manage complex, and potentially confusing, microworlds CPS tasks. We do this by eliciting the role of non-cognitive variables on people's capacity to deal with the task as they gain more experience with it but also as it becomes more complex. Based on the literature reviewed, we have two sets of hypotheses about the influence of cognitive and non-cognitive variables on performance – the first based on stage performance-*levels* (i.e., stage means), and the second on stage performance *trajectories* (ie., stage slopes).

Hypothesis Set 1: Performance-Level Effects

Our first set of hypotheses addresses how cognitive ability and non-cognitive traits influence mean differences between task complexity stages[1]. First, we expect a complexity-by-general mental ability (GMA) effect – that is, mean stage performance will decline concomitant with increases in task complexity, but that those with higher GMA will be less impacted (Hypothesis 1). Second, we hypothesise that several intellectual investment traits and personality facets and domains will lessen the mean performance decrease as stage complexity increases and that these effects will persist above and beyond GMA. These non-cognitive traits are all associated with motivation to explore, engage with, and master novel cognitive tasks. We expect mean stage performance for those higher in NFC and Learning Goal Orientation will decline less substantially as stage complexity increases because these traits are likely to facilitate curiosity, exploration, and learning of this novel and challenging system (Hypotheses 2A and 2B). We also expect the same effect for those high in Intellect, the Intellect-Conquer process, and the Intellect-Learning and Intellect-Thinking operations for the same reasons (Hypothesis 2C). Finally, we hypothesise that those higher in Open-Mindedness and its Intellectual Curiosity and Creative Imagination domains (Hypothesis 3A), and Conscientiousness and its Productiveness domain (Hypothesis 3B) will demonstrate smaller performance declines over time for the same reasons. Finally, and separately to the hypotheses above, we expect mean performance for those higher in Negative Emotionality and its Emotional Volatility and Anxiety domains (Hypothesis 3C) to decline more dramatically over time than those lower in these traits

[1] In our analyses, mean performance represents average performance where trial number is equal to zero. As trial number is centred overall and within stages of the task in our analysis, in practical terms, mean performance refers to average performance at the midpoint of the task or stage. When examining contrasts between stages, mean performance differences represent the difference in estimated penalty score at the midpoint of one stage relative to the estimated penalty score at the midpoint of the contrasted stage.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

because they are likely to find this challenging and uncertain cognitive task more difficult to manage.

Hypothesis Set 2: Performance-Trajectories Effects

Our second set of hypotheses addresses how cognitive ability and non-cognitive traits influence the differences in within-stage performance trajectories between complexity stages. Within-stage performance trajectories provide an indication of how people's performance changes throughout the stage and, by extension, how well they learn to control the microworlds system at a given stage. First, we expect a complexity-by-general mental ability (GMA) effect such that within-stage trajectories will be steeper concomitant with increases in task complexity as people find the task more challenging to learn to control, but that the trajectories of those with higher GMA will be less impacted as complexity increases (Hypothesis 4). We expect several non-cognitive traits to influence slopes above and beyond GMA. We hypothesise that those high in NFC and Learning Goal Orientation will exhibit shallower slopes with increasing complexity as they will learn to control complexity increases more effectively than those lower in NFC and Learning Goal Orientation (Hypotheses 5A and 5B). We expect the same effects for those high in Intellect, specifically the Intellect-Conquer process and the Intellect-Learning and Intellect-Thinking operations (Hypotheses 5C), for those high in Open-Mindedness and its Intellectual Curiosity and Creative Imagination domains (Hypothesis 6A), and Conscientiousness and its Productiveness domain (Hypothesis 6B). Finally, we expect that performance trajectory slopes for people high in Negative Emotionality and its Emotional Volatility and Anxiety domains will be steeper as complexity increases, as these traits will have a negative influence on the capacity to learn to control these confusing and challenging complexity increases (Hypothesis 6C) (Soto & John, 2017).

We make no specific hypotheses about Agreeableness and Extraversion but will conduct exploratory analysis of these and other trait and domain-level mean and slope associations.

2. Materials and Methods

Participants

Participants were recruited as part of a larger study via M-Turk (n = 252) (see Fayn et al. (2019) for further details). Participants were paid $10 USD for their participation, around an hour of time. All participants reported living in the USA and had completed at least 100 approved M-Turk studies with a 95% approval rating or above. Nineteen participants were excluded for one of the following reasons: failing attention checks (n = 7), failing the practice task (n = 5), or non-serious attempts on one or more tasks in the study (n = 7). This left a final sample of n = 233 participants (n_{Male} = 117, n_{Female} = 115, $n_{Unspecified}$ = 1). The average age of participants in the final sample was 34.99 years (SD = 10.03).

Materials

CPS Microworld

The CPS task was a microworld simulation based on the inventory management simulation used in Birney et al. (2018). The broad design of Birney et al.'s (2018) task was maintained, however, complexity manipulations were introduced in one block rather than across blocks, and the context was adapted for a general audience. Participants were told that they had been put in charge of successfully managing the ecosystem of an island, and to do so, had to maintain the number of species on the island at 100 species. To achieve this goal, each trial they were asked to make a decision to introduce new species or relocate species back to the mainland. This decision was made by entering a positive number (introduce species) or negative number (relocate species) into the box and observing the subsequent effects on the number of species. This is illustrated in Figure 1.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

Figure 1. *Microworld Simulation Interface After Decisions*

Participants were also informed that the system was volatile and that the relationship between decisions and outcomes could change throughout the course of the task. Two aspects of the microworld were manipulated to increase the complexity of the task – outflows and delays. Outflows were the loss of species that occurred at each trial independent of the participants' decision and could be fixed and constant (e.g. 20 species lost per trial) or random (e.g. random number between 10-30 species lost per trial). Delays were the trial lag between decisions being made and decisions taking effect. In this simulation, participants experienced three variants of outflow/delay manipulations that demarcate the three stages within the single block of the task:

- Stage 1 – low complexity: From trials 1-23 there was a constant outflow. In trials 1-11 the constant outflow was 10 species per trial, and in trials 12-23 the constant outflow was 20 species per trial.
- Stage 2 – medium complexity: From trials 24-42 there was a random outflow of between 10 and 30 species per trial (average of 20 species).
- Stage 3 – high complexity: From trials 43-61 there was a random outflow of between 10 and 30 species per trial and a delay of two trials between decisions being made and decisions taking effect.

Based on the results from Birney et al. (2018), we expected that these three stages of the task would be of increasing complexity for participants but that this increase would not necessarily be linear in its effect on performance.

The task also included two "mass extinction events" on trials 6 and 42, where 50 species were lost on a single trial. These extinction events were included to replicate the dynamic and complex nature of real-world complex systems that can spontaneously change. The second extinction event on trial 42 also encouraged participants to notice the change in the delay rule; without the extinction event, the participant could theoretically have maintained the same strategy from the second stage.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

Participants were also asked to report on their emotional state before the microworlds task began, every three trials during the task, and at the conclusion of this task, however, this data was not used for the purposes of this study.

Intellectual Investment Trait Measures

NFC. NFC was assessed using the 18-item scale developed by Cacioppo and Petty (1982) and had excellent Cronbach's α reliability of α = 0.95.

Goal Orientation. Goal Orientation was assessed using the 16-item scale developed by VandeWalle (1997). The overall goal orientation scale had acceptable Cronbach's α reliability of α = 0.75. The 6-item Learning sub-scale had excellent Cronbach's α reliability of α = 0.92, the 5-item Performance-Avoid sub-scale had excellent Cronbach's α reliability of α = 0.91, and the 5-item Performance-Prove sub-scale had acceptable Cronbach's α reliability of α = 0.76.

Intellect. Intellect was assessed using the 24-item item scale developed by Mussel (2013) and had excellent Cronbach's α reliability of α = 0.98. The reliability for all facets of the Intellect scale was also excellent (Intellect Seek α = 0.96; Intellect Conquer α = 0.96; Intellect Think α = 0.95; Intellect Learn α = 0.94; Intellect Create α = 0.94).

Personality

Personality was assessed using the 60-item Big Five Inventory-2 (BFI-2) (Soto & John, 2017). The BFI-2 is a freely available measure of the Five Factor Model of personality with high internal consistency and test-retest reliability (Soto & John, 2017). The BFI-2 assesses five personality trait domains and fifteen facets – three per domain. Each domain has one facet that most strongly loads on the domain factor: Extraversion-Sociability, Agreeableness-Compassion, Conscientiousness-Organisation, Negative Emotionality-Anxiety/Fear, and Open Mindedness-Intellectual Curiosity. All domains had good to excellent Cronbach's α reliability (Extraversion α = 0.92; Agreeableness α = 0.88; Conscientiousness α = 0.92; Negative Emotionality α = 0.94; Open-Mindedness α = 0.91). All facets had good to excellent Cronbach's α reliability (Extraversion: Sociability α = 0.92; Extraversion: Assertiveness α = 0.87; Extraversion: Energy Level α = 0.82; Agreeableness: Compassion α = 0.81; Agreeableness: Respectfulness α = 0.78; Agreeableness: Trust α = 0.81; Conscientiousness: Organisation α = 0.87; Conscientiousness: Productiveness α = 0.88; Conscientiousness: Responsibility α = 0.78; Negative Emotionality: Anxiety α = 0.85; Negative Emotionality: Depression α = 0.90; Negative Emotionality: Emotional Volatility α = 0.89; Open-Mindedness: Intellectual Curiosity α = 0.82; Open-Mindedness: Aesthetic Sensitivity α = 0.86; Open-Mindedness: Creative Imagination α = 0.83).

General Mental Ability (GMA)

GMA was assessed using a 16-item version of the International Cognitive Ability Resource (ICAR) (Condon & Revelle, 2014). The ICAR is an open-access cognitive ability assessment that is strongly correlated with longer and more complex measures of GMA (r=.81 with full-scale IQ and r=.94 with a CFA-derived latent g factor in Young & Keith, 2020). This version of the ICAR included four items each for verbal reasoning, letter-number series, matrix reasoning, and 3D object rotation. ICAR score was derived as the total sum score of all correct items. The full ICAR had acceptable Cronbach's α reliability of α = 0.75.

Procedure

Participants provided consent to participate and then completed the NFC and Intellect scales, and the BFI-2. Then, they were given detailed instructions on the microworld and were asked to answer three questions about the instructions to ensure they had read and understood the requirements. All three questions had to be answered correctly to advance to the microworld. Participants were given three attempts to answer all three questions correctly. Following this, participants were given a practice block of 10 microworld trials. The practice block followed the easiest rule variant – a constant outflow of 5 species per trial. Halfway through the practice block, participants were informed of this rule and told to continue the final 5 practice trials applying this rule. Following the practice block, participants completed the 61-trial microworld task, the Goal Orientation scale, and the ICAR in that order, and then given a debrief.

3. Results

Preliminary Analyses

For each trial participants received a penalty score, which was calculated as the deviation between the number of species after their decision was made and the ideal number of species (100). All scores for intellectual investment traits, personality traits, and cognitive ability were transformed to z-scores. Analysis was conducted using SPSS v26.0 (IBMCorp, 2019) and R version 3.6.2 (R Core Team, 2019). Generalized linear mixed-effects models (GLMM) were estimated using the R package lme4 (Bates et al., 2015). For all GLMM we treated penalty scores as a count variable modelled using a negative binomial distribution. This decision was made because penalty scores could take positive, integer values only thus penalty score is a discrete rather than continuous variable. Furthermore, analysis of the dispersion statistics showed that variance in penalty score was substantially greater than the mean penalty score, overall and for each stage of the task, violating the assumptions of the poisson distribution (roughly equal variance and mean for the dependent variable) but fitting with the assumptions of the negative binomial distribution. For the remaining of the paper, all variables are demarcated as such using italicised font.

Descriptive Statistics

Correlations between mean penalty score overall and at each stage of the task with domain level variables are reported in Table 1 (the full matrix of domain inter-correlations is contained in the Supplementary Materials). Only *Learning Goal Orientation* and *NFC* were significantly related to overall mean penalty score. As expected, there was some evidence for a change in the nature of the performance correlations across the three stages. *Conscientiousness* was positively related to performance in the first two stages but negatively related in the final stage. Moreover, the associations between performance and *NFC* and *Learning Goal Orientation* emerged only in the final and most complex stage of the task. The association between performance and *GMA* declined in magnitude from the first to the second stage, and was not significant in the final high complexity stage.

At the facet level and across all stages, the associations that emerged were few in number and small in size (see Table 2, full inter-correlation matrix is in the Supplementary Materials). Patterns of substantive interest are, however, apparent. For instance, *Extraversion* facets tended to have positive correlations in Stages 1 and 2 but negative or zero correlations in the last stage. A number of the Intellect facets were also significant predictors for performance, but only in the final high complexity stage. Finally, for *Conscientiousness*, the *Organisation* facet predicts worse performance only in the low complexity stage whereas the *Productiveness* facet predicts better performance only in the high complexity stage; we note that these effects are in the opposite direction. Thus, consistent with the trait-level factors, there is some limited evidence for a change in the nature and direction of performance correlations across stages as complexity increases.

Table 1. *Domain Correlations with Mean Penalty Score Overall and by Stage Complexity*

	Overall	Stage 1 Low	Stage 2 Medium	Stage 3 High
GMA	-.055	**-.308**	**-.199**	.066
Open-Mindedness (O)	-.025	.092	-.011	-.049
Conscientiousness (C)	-.066	**.175**	**.159**	**-.178**
Extraversion (E)	-.094	.065	-.004	-.048
Agreeableness (A)	-.030	.126	.086	-.120
Negative Emotionality (N)	.108	-.013	-.001	.115
NFC	**-.156**	-.008	-.009	**-.157**
Intellect	-.109	.028	.010	-.122
Learning Goal Orientation (LGO)	**-.184**	-.051	-.026	**-.171**
Performance-Prove Goal Orientation (PGO)	-.078	-.065	.056	-.074
Performance-Avoid Goal Orientation (AGO)	.084	-.047	.006	.099

Correlations in **bold** are significant at $\alpha = .05$. Stage 1 = trials 1 to 23; Stage 2 = trials 24 to 42; Stage 3 = trials 43-61.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

Table 2. *Facet Correlations with Mean Penalty Score Overall and by Stage Complexity*

	Overall	Stage 1 Low	Stage 2 Medium	Stage 3 High
O - Aesthetic Sensitivity (O-AesSens)	.021	.110	.021	-.013
O - Intellectual Curiosity (O-IntCur)	-.029	.026	-.062	-.025
O - Creative Imagination (O-CreImag)	-.063	.094	.006	-.092
C - Organisation (C-Org)	.021	**.144**	.116	-.041
C - Productiveness (C-Prod)	**-.121**	.114	.098	**-.176**
C - Responsibility (C-Resp)	-.078	.066	.001	-.099
E - Sociability (E-Soc)	-.101	**.171**	**.173**	**-.186**
E - Assertiveness (E-Asrt)	-.107	.112	**.136**	**-.169**
E - Energy Level (E-EnLev)	-.028	**.167**	.094	-.094
A - Compassion (A-Comp)	-.012	.089	.000	-.037
A - Respectfulness (A-Resp)	-.064	-.003	-.029	-.059
A - Trust (A-Tr)	-.054	.029	-.008	-.062
N - Anxiety (N-Anx)	.095	-.022	-.008	.106
N - Depression (N-Dep)	.088	-.041	-.039	.110
N - Emotional Volatility (N-EmVol)	**.111**	.032	.050	.096
Intellect - Seek (Int-Sk)	**-.128**	.007	.000	**-.134**
Intellect - Conquer (Int-Cq)	-.082	.048	.021	-.102
Intellect - Think (Int-Tk)	-.089	.004	.002	-.094
Intellect - Learn (Int-Lrn)	-.098	.019	-.008	-.105
Intellect - Create (Int-Crt)	**-.127**	.056	.033	**-.154**
Intellect - SeekThink (Int-SkTk)	-.093	-.031	-.016	-.085
Intellect - SeekLearn (Int-SkLrn)	**-.111**	-.009	-.040	-.104
Intellect - SeekCreate (Int-SkCrt)	**-.154**	.054	.046	**-.184**
Intellect - ConquerThink (Int-CqTk)	-.078	.042	.022	-.097
Intellect - ConquerLearn (Int-CqLrn)	-.076	.044	.023	-.095
Intellect - ConquerCreate (Int-CqCrt)	-.084	.052	.015	-.104

Correlations in **bold** are significant at $\alpha = .05$. Stage 1 = trials 1 to 23; Stage 2 = trials 24 to 42; Stage 3 = trials 43-61.

Analyses Overview

For the GLMM analyses, we report incidence rate ratios (IRR) and 95% confidence intervals for the IRR to three decimal places. The IRR reflects the expected factor change in the dependent variable, penalty score, for a one standard deviation increase in the independent variable, non-cognitive trait score (because the non-cognitive variables have been standardised). Trial number was mean-centred (1) over all 61 trials and (2) within stages, which we refer to as *total-trial* and *stage-trial*, respectively. Stage was coded as three levels, each representing a different rule complexity—low, medium, and high—as outlined in the methods. We created two orthogonal effect-coded complexity contrasts to model how non-cognitive variables differentially influenced participants' capacity to deal with these complexity changes across stages: (1) *medium-demand* complexity contrast effect (Stage 1 versus Stage 2, representing the capacity to manage low complexity versus medium complexity), and (2) *high-demand* complexity contrast effect (Stage 1 and 2 on average versus Stage 3, representing the capacity to manage low and medium complexity versus high complexity). We ran one model testing the influence of *GMA* on performance, then ran a series of additional models examining the role of each non-cognitive trait (at the domain and facet levels) on performance (with and without *GMA*). In each non-cognitive moderation model, we examine four output coefficients:

1. The two-way interaction between the non-cognitive variable and the *medium-demand* complexity contrast. This is a test of whether mean performance differs between the low and medium complexity stages as a function of the non-cognitive variable and is linked to Hypothesis Set 1.
2. The two-way interaction between the non-cognitive variable and the *high-demand* complexity contrast. This is a test of whether mean performance differs between the low and medium complexity stages and the high complexity stage as a function of the non-cognitive variable and is linked to Hypothesis Set 1.
3. The three-way interaction between the non-cognitive variable, the *medium-complexity* contrast, and stage-trial number. This is a test of the strength and direction of differences in within-stage performance trajectory slopes between the low and medium complexity stages as a function of non-cognitive traits and is linked to Hypothesis Set 2.
4. The three-way interaction between the non-cognitive variable, the *high-complexity* contrast, and stage-trial number. This is a test of the strength and direction of differences in within-stage performance trajectory slopes between the low and medium complexity stages and the high complexity stage as a function of non-cognitive traits and is linked to Hypothesis Set 2.

These complexity contrasts provided additional information on which to base our conclusions rather than relying on correlational analyses alone. Full specifications of the models run can be found in the supplementary materials.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

Overall Findings

To begin, the raw penalty score data is plot by trial number in Figure 2(a). We ran a basic model on this data (as described above) testing whether mean performance differed by stage complexity (see Model 1, supplementary materials). As expected, there was a significant *medium-demand* complexity effect (*IRR* = 1.149, *95% CI* = 1.101 – 1.200, $p < .001$), and *high-demand* complexity effect (*IRR* = 8.134, *95% CI* = 7.837 – 8.442, $p < .001$). Mean penalty score was 1.149 times higher in the medium complexity stage than the low complexity stage, suggesting participants found random outflows more challenging to manage than constant outflows. Mean penalty score in the high complexity stage was 8.134 times higher than the low and medium complexity stages on average, suggesting participants found delays substantially more challenging to manage than no delays. There was also a *high-demand* complexity effect on performance trajectory slopes (*IRR* = 1.052, *95% CI* = 1.045 – 1.058, $p < .001$). Specifically, the penalty score trajectory under high complexity was 1.052 times steeper than the average penalty score trajectory under low and medium complexity conditions, suggesting delays impeded people's capacity to learn rules and control the system. There was no *medium-demand* complexity effect on performance trajectory slopes, suggesting the introduction of random outflows did not affect people's capacity to learn rules and control the system. Therefore, as hypothesised and in line with Birney et al. (2018), increases in complexity throughout the task were related to performance decreases (increased penalty score). The results from this model are illustrated in Figure 2b, which shows that average penalty score increased somewhat exponentially throughout the task.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

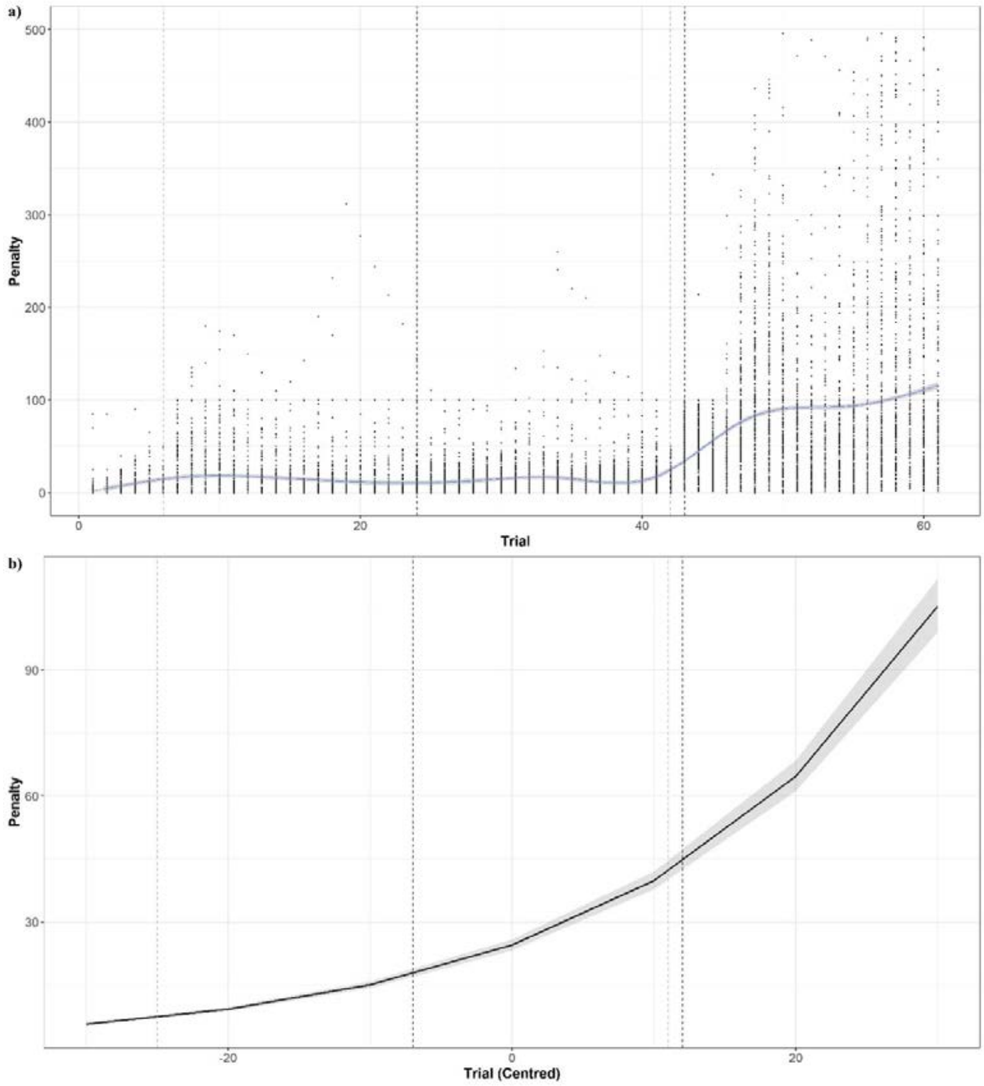

Figure 2. *Plots of (a) Raw and Smoothed, (b) GLMM Modelled Penalty Scores Over Time (Model 1). Note: Stage transitions are demarcated by black dashed lines. Mass extinction events are demarcated by grey dashed lines. Trial number has been centred for the GLMM plot and preserved in this figure to appropriately represent the underlying model.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

General Mental Ability

Model 2 tested the interaction between *GMA* and total-trial number on overall performance, which is represented in Figure 3. The main effect for *GMA* was statistically significant as expected. Those of higher *GMA* tended to have lower penalty scores (*IRR* = 0.884, *95% CI* = 0.842 – 0.928, $p < .001$). Model 3 tested interactions between GMA, the complexity contrasts, and stage-trial number to understand how *GMA* differentially influenced performance in different stages. First, *GMA* was associated with significant mean performance differences for the *medium-demand* complexity contrast (*IRR* = 1.152, *95% CI* = 1.103 – 1.202, $p < .001$) and *high-demand* complexity contrast (*IRR* = 1.241, *95% CI* = 1.196 – 1.288, $p < .001$). Both effects were in the opposite direction to the main effect of GMA on performance in Model 2. Second, there were significant interactions effects of *GMA* on performance trajectory slopes for the *medium-demand* complexity contrast (*IRR* = 1.021, *95% CI* = 1.014 – 1.029, $p < .001$) and *high-demand* complexity contrast (*IRR* = 1.009, *95% CI* = 1.003 – 1.016, $p = .006$). Taken together with the mean performance findings, this implies that the influence of *GMA* on performance changed throughout the task. Indeed, as illustrated in Figure 3, *GMA* did not seem to differentially predict trajectories in the first stage but did in the second and third stages. In the medium complexity stage, trajectories by *GMA* level narrowed over time and in the high complexity stage, trajectories by *GMA* level crossed over time. This suggests that all else equal, those high in *GMA* were not necessarily better equipped to deal with complexity increases over time, despite performing better overall in the first stage. Therefore, Hypothesis 1 and Hypothesis 4 were not supported.

A series of separate analyses examining the additional predictive utility of each of the non-cognitive variables as covariates to Model 3 did not statistically improve model fit.

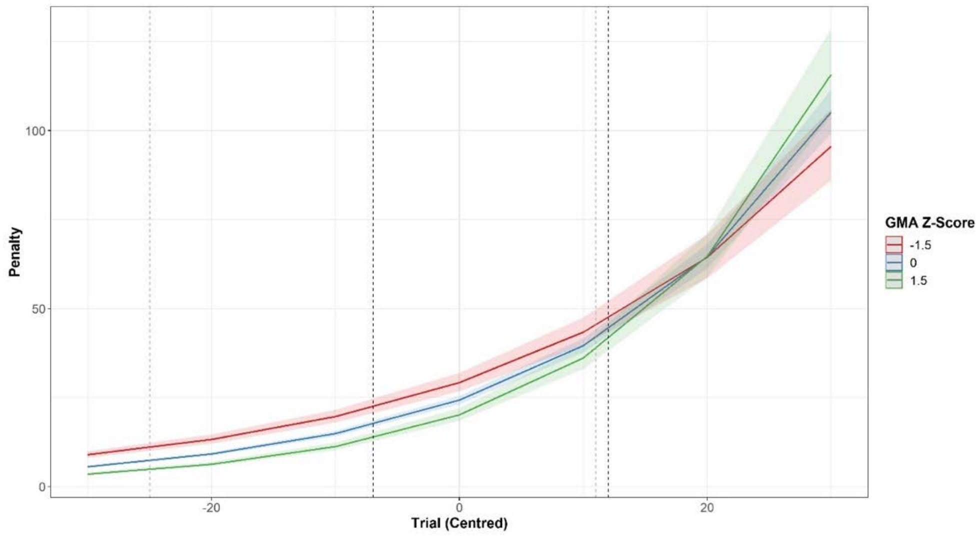

Figure 3. *Performance Trajectory by GMA Scores (Mean ± 1.5 SD) from Model 2. Note: Stage transitions are demarcated by black dashed lines. Mass extinction events are demarcated by grey dashed lines. Trial number has been centred and preserved in this figure to appropriately represent the underlying model.*

Non-Cognitive Moderation Effects

To investigate the influence of non-cognitive traits on performance trajectories, the intellectual investment and personality traits were standardised, and the stage and trial variables were effect coded as described above. Model 4 represents the specification of the generic interaction analysis. This was run for each non-cognitive trait moderator separately. The reason for analysing the traits separately, rather than in one combined analysis, is two-fold. First and foremost, we are interested in overall effect of a moderator conditional (or not) on *GMA*. We are not concerned here with whether the moderator effects hold conditional on the range of other traits we have assessed. Second, the effect sizes reported for moderators are small and additional covariates would serve to reduce statistical power. We also recognise many separate analyses are likely to raise questions about Type 1 error rates. As outlined by Birney et al. (2017), the multi-level approach used is distinctly advantageous in this regard (Gelman et al., 2012) compared to OLS regression (Brunner & Austin, 2009). It uses a partial pooling process (often referred to as "shrinkage") that serves to shift parameter estimates and their associated standard errors toward mean coefficients in the complete data. This processes has the desirable effect of shrinking coefficients that are estimated with small accuracy more so than those estimated with higher accuracy (Hox, 2010), thus intervals for comparisons are more likely to include zero (Gelman et al., 2012).

Analyses were run first without (Model 4A) and then with *GMA* (Model 4B) to test whether the effects of intellectual investment or personality traits remained when *GMA* was controlled for. We first report the domain-level results and then review the facet-level results, which are summarised across Figures 4 – 7. Figure 4a contains the domain-level two-way interaction effects between each non-cognitive trait and the *medium-demand* contrast, and Figure 4b with the *high-demand* contrast. Figure 5 contains the domain-level three-way interaction effects from the same model, but also includes within-stage trial order as a variable (i.e., the effects test the strength and direction of (1) between-stage differences in (2) performance trajectories as a function of (3) non-cognitive traits). Figure 5a reports these effects for the *medium-demand* contrast, and Figure 5b the *high-demand* contrast. Analogously, Figures 6a and 6b contains the two-way interactions, and Figures 7a and 7b contains the three-way interactions for the facet-level moderators for both complexity contrasts.

Domain Level Effects: In Hypothesis Set 1, we hypothesised that *NFC* (Hypothesis 2A), *Learning Goal Orientation* (Hypothesis 2B), *Intellect* (Hypothesis 2C), *Open-Mindedness* (Hypothesis 3A), *Conscientiousness* (Hypothesis 3B), and *Negative Emotionality* (Hypothesis 3B) would be the strongest trait drivers of mean performance differences between stages. Figure 4 shows that, for most domains, the relationship between non-cognitive traits and mean performance differed between the complexity stages. Five of the ten domains—*Perform-Prove Goal Orientation*, *Agreeableness*, *Conscientiousness*, *Extraversion*, and *Open-Mindedness*—were associated with *medium-demand* contrast effects, and the inclusion of *GMA* did not alter the statistical significance of these effects. All ten domains except *Perform-Prove Goal Orientation* were associated with *high-demand* contrast effects, which all remained significant when controlling for *GMA*. Our hypotheses were therefore supported for *Open-Mindedness* (Hypothesis 3A) and *Conscientiousness* (Hypothesis 3B), which were associated with both *medium-demand* and *high-demand* mean performance differences. Our hypotheses were partially supported for *NFC* (Hypothesis 2A), *Learning Goal Orientation* (Hypothesis 2B), *Intellect* (Hypothesis 2C), and *Negative Emotionality* (Hypothesis 3C), which were associated with mean performance differences but only for the *high-demand* contrast.

Few of the hypothesised trajectory slope effects from Hypothesis Set 2 were supported, as shown in Figure 5. We hypothesised that those high in *NFC* (Hypothesis 5A),

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

Learning Goal Orientation (Hypothesis 5B), *Intellect* (Hypothesis 5C), *Open-Mindedness* (Hypothesis 6A), and *Conscientiousness* (Hypothesis 6B) would better manage complexity increases over time, however, we only found mixed evidence to support Hypotheses 5C and 6B. Higher *Intellect* seemed to confer benefits in managing the *high-demand* complexity shift when controlling for *GMA*. High *Conscientiousness* also seemed to confer benefits in managing the *medium-demand* complexity shift controlling for *GMA*. We also hypothesised that high *Negative Emotionality* would be associated with sharper performance decreases (i.e., steeped performance trajectories) as complexity increased (Hypothesis 6C) but the evidence for this hypothesis was also mixed. High *Negative Emotionality* did tend to exacerbate the *high-demand* complexity shift, but this effect did not persist when controlling for *GMA*.

In addition, we uncovered several unexpected domain level effects in our exploratory analysis. High *Extraversion* was associated with improved capacity to manage the *medium-demand* complexity shift, which persisted when controlling for *GMA*. High *Performance-Prove Goal Orientation* had the same effect on the capacity to manage the *high-demand* complexity shift, which remained when controlling for *GMA*. Figures 8 and 9 illustrate how these mean performance and slope differences manifest in performance trajectories for two example domains: *Conscientiousness* and *Extraversion* (controlling for *GMA*)[2].

[2] For illustrative purposes, Figures 8 and 9 are representations of analyses that included total-trial number and the respective moderator, controlling for GMA. The respective R models are:

Figure 8: Penalty = ICAR.z + *trial.c*Conscientiousness.c* + (1|ID);

Figure 9: Penalty = ICAR.z + *trial.c*Extraversion.c* + (1|ID)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

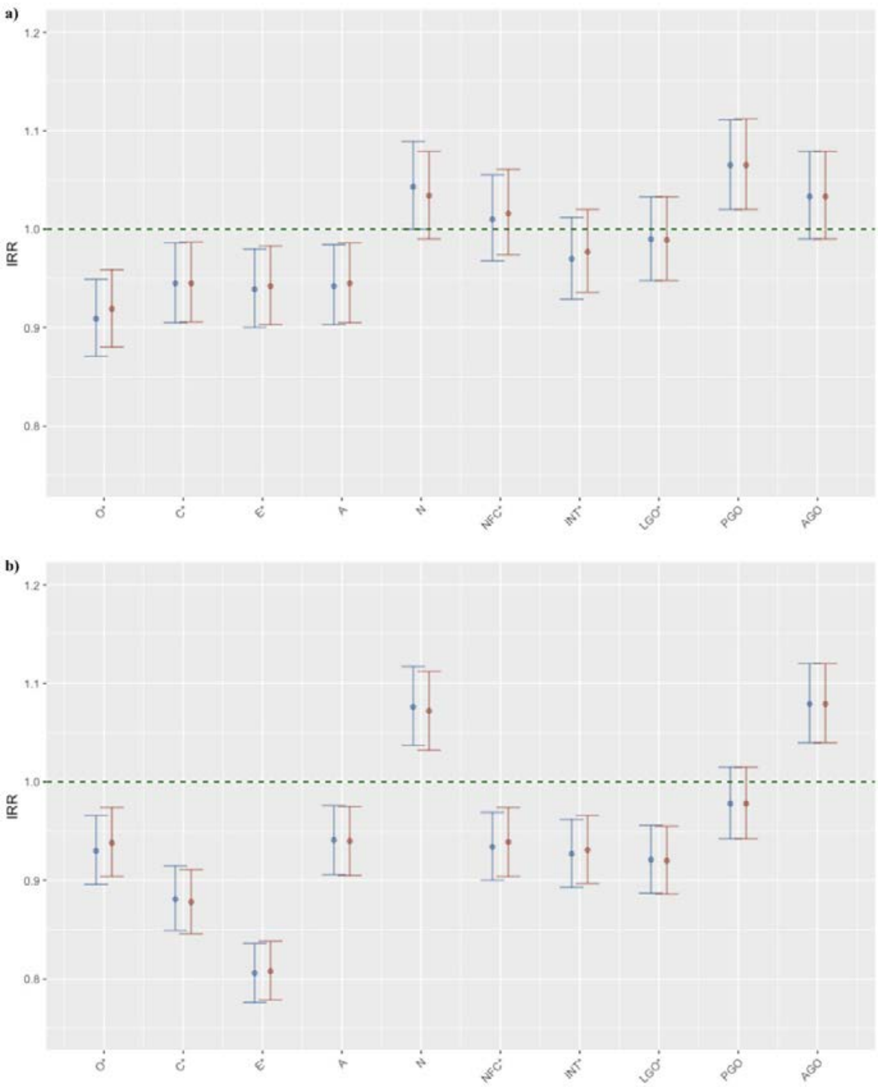

Figure 4. *Mean performance by domain interaction effects for a) medium-demand and b) high-demand complexity contrasts. Note: Blue bars indicate relations from Model 4a; Red bars indicate relations from Model 4b. The variable key is provided in Table 2. Variables with asterisks (*) are those for which we made specific hypotheses.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

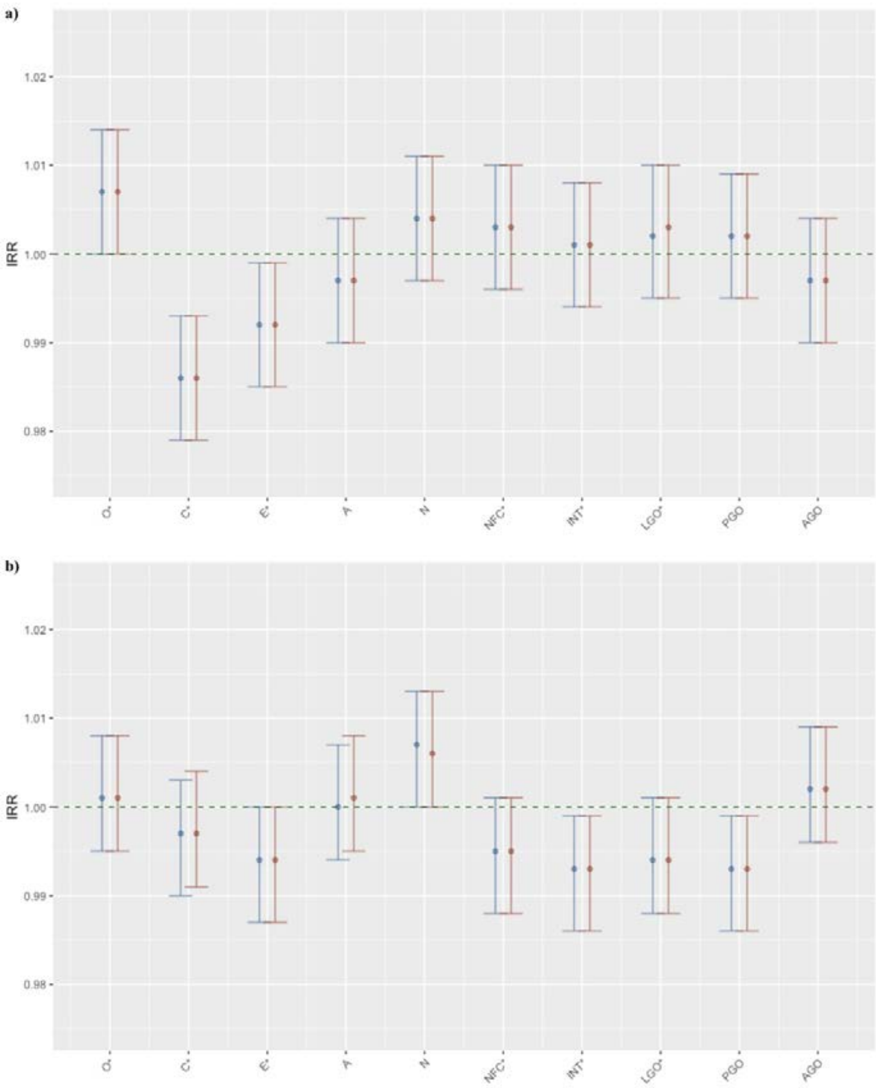

Figure 5. *Performance trajectory slopes by domain interaction effects for a) medium-demand and b) high-demand complexity contrasts. Note: Blue bars indicate relations from Model 4a; Red bars indicate relations from Model 4b. The variable key is provided in Table 2. Variables with asterisks (*) are those for which we made specific hypotheses.*

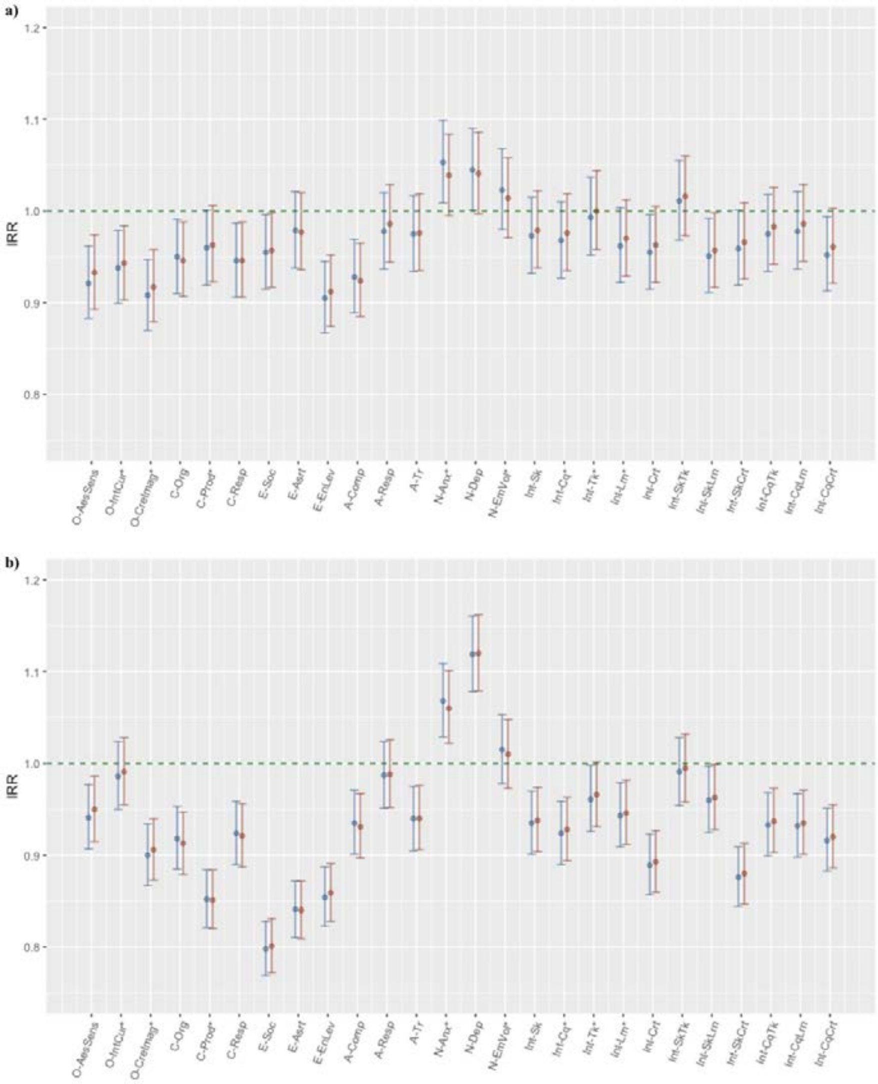

Figure 6. *Mean performance by facet interactions for a) medium-demand and b) high-demand complexity contrasts. Note: Blue bars indicate relations from Model 4a; Red bars indicate relations from Model 4b. The variable key is provided in Table 2. Variables with asterisks (*) are those for which we made specific hypotheses.*

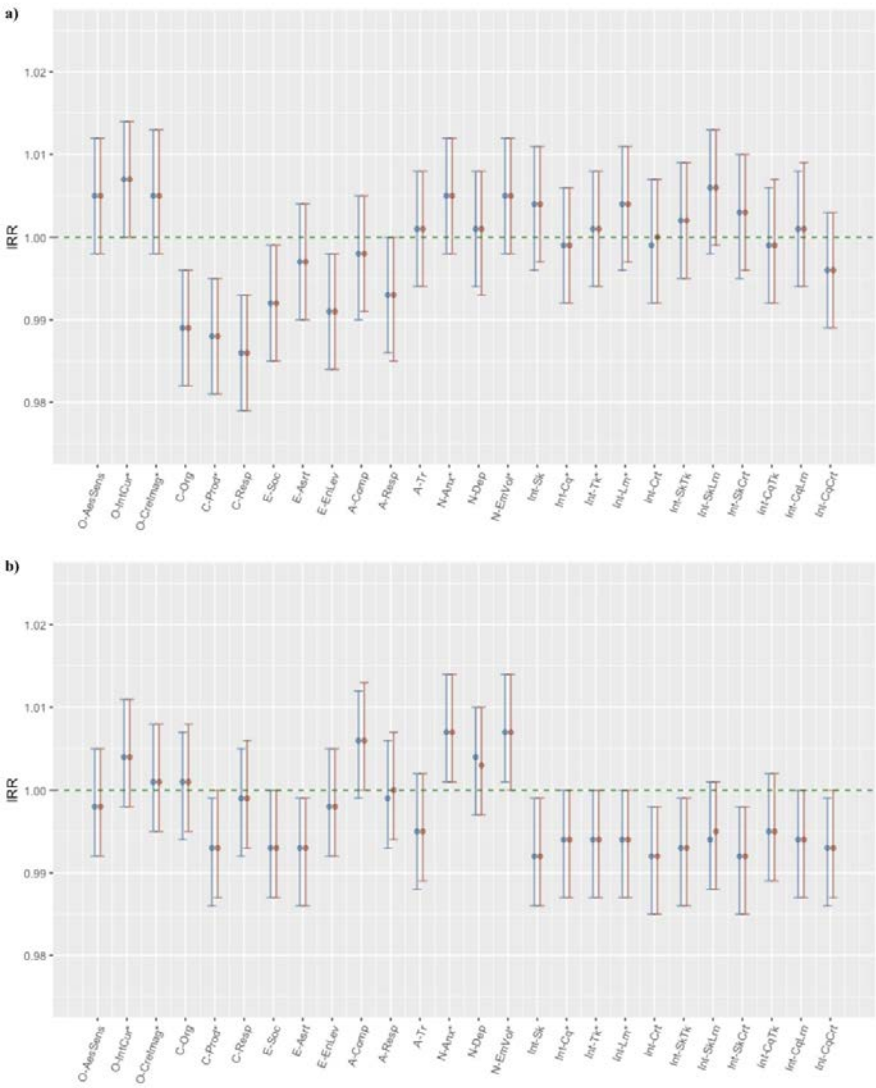

Figure 7. *Performance trajectory slopes by facet interactions for a) medium-demand and b) high-demand complexity contrasts***.** *Note: Blue bars indicate relations from Model 4a; Red bars indicate relations from Model 4b. The variable key is provided in Table 2. Variables with asterisks (*) are those for which we made specific hypotheses.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

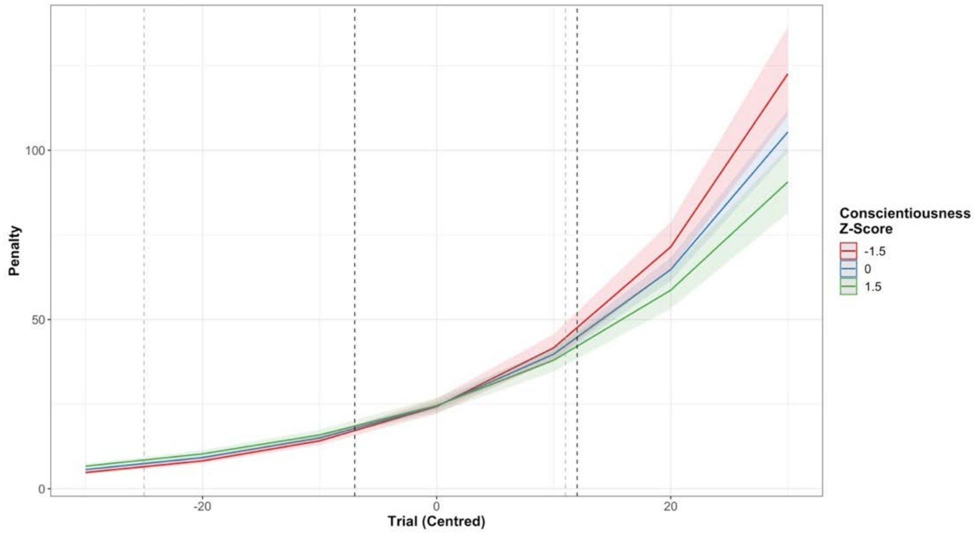

Figure 8. *Performance trajectory by Conscientiousness (controlling for GMA). Note: Stage transitions are demarcated by black dashed lines. Mass extinction events are demarcated by grey dashed lines. Trial number has been centred and preserved in this figure to appropriately represent the underlying model.*

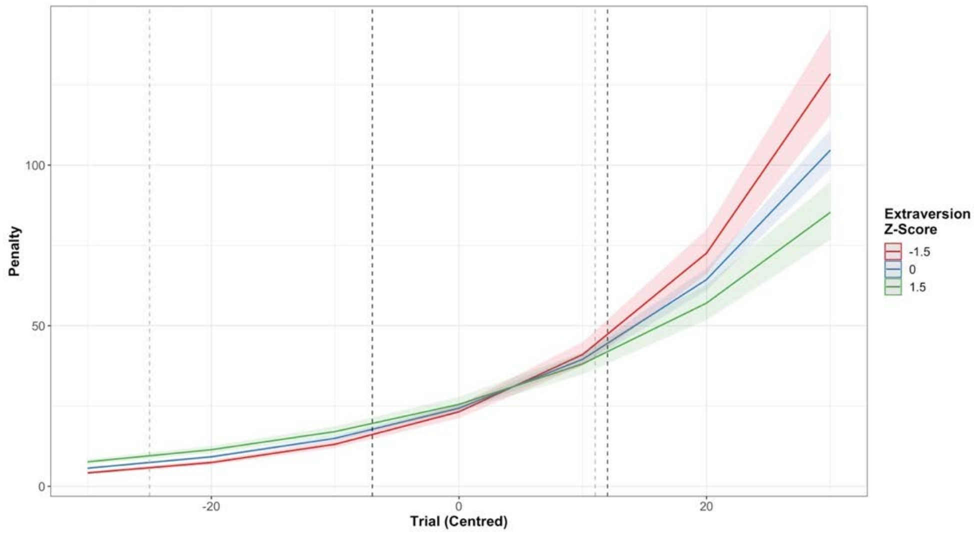

Figure 9. *Performance trajectory by Extraversion (controlling for GMA). Note: Stage transitions are demarcated by black dashed lines. Mass extinction events are demarcated by grey dashed lines. Trial number has been centred and preserved in this figure to appropriately represent the underlying model.*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

Facet Level Effects: Figure 6a demonstrates that 11 of the 26 facets were associated with *medium-demand* mean performance differences, however, only nine of these effects remained statistically significant after controlling for *GMA*. Figure 6b shows that 22 of the 26 facets were associated with *high-demand* mean performance differences, however, only 21 of these effects remained statistically significant after controlling for *GMA*. Figure 7a contains the facet level slope effects; five of the 26 traits were associated with *medium-demand* slope differences and a further five of the 26 were associated with *high-demand* slope differences (Figure 7b). Given the number of facet level effects, we first focus on those we made specific hypotheses about, and then others of theoretical and statistical significance. Furthermore, we note that the Intellect facets were all highly correlated with each other (r = .79-.97) and therefore the extent to which they offer unique substantive information in explaining variation in performance is low. We include the variables in our analyses as they were planned from the outset but advocate caution in interpreting the results for these variables as a result.

We hypothesised that those high in *Intellect Conquer* process and *Intellect Learning* operation (Hypothesis 2C), *Open-Mindedness: Intellectual Curiosity* and *Creative Imagination* (Hypothesis 3A), and *Conscientiousness: Productiveness* (Hypothesis 3B) would deal better with complexity increases, above and beyond GMA. Hypothesis 3A was supported for *medium-demand* mean performance differences. Hypotheses 2C and 3B were supported for *high-demand* mean performance differences. Hypothesis 3A was partially supported for *high-demand* mean performance differences, with only *Open-Mindedness: Creative Imagination* being associated with *high-demand* mean performance differences. We also hypothesised that those with low *Negative Emotionality: Emotional Volatility* and *Anxiety* would be better equipped to manage complexity increases (Hypothesis 3C). This hypothesis was only supported for one facet, *Negative Emotionality: Anxiety*, for the *high-demand* mean performance difference.

We also hypothesised that those high in *Intellect Conquer* process and *Intellect Learning* operation (Hypothesis 5C), *Open-Mindedness: Intellectual Curiosity* and *Creative Imagination* (Hypothesis 5A), and *Conscientiousness: Productiveness* (Hypothesis 5B) would deal better with complexity increases as indicated by shallow performance trajectory slopes. Figure 7 contains the facet level slope effects and shows that the only hypothesis supported was Hypothesis 5B, *Conscientiousness: Productiveness*, which conferred performance benefits in managing the *medium-demand* and *high-demand* shifts, controlling for *GMA*. We also hypothesised that those with low *Negative Emotionality: Emotional Volatility* and *Anxiety* would be better equipped to manage complexity increases (Hypothesis and 6C). This effect was supported for both facets but only for the *high-demand* complexity.

Echoing the domain level results, we also observed some consistent and unexpected effects of various facets on the capacity to manage complexity changes. In addition to *Conscientiousness: Productiveness*, both high *Conscientiousness: Organisation* and high *Conscientiousness: Responsibility* conferred mean performance and shallow learning slope benefits in managing the *medium-demand* shift, which persisted when controlling for *GMA*. These variables did not, however, have the same effect for the *high-demand* complexity shift for either mean performance or slope differences. As may be expected from the domain level findings, the extraversion facets had a significant effect on both complexity changes. Higher *Extraversion: Sociability* and *Extraversion: Energy Level* facilitated improved management of the *medium-demand* complexity shift (mean performance and slopes) when controlling for *GMA*. Also, higher *Extraversion: Sociability* and *Extraversion: Assertiveness Level* facilitated improved management of the *high-demand* complexity shift (mean performance and slopes) when controlling for *GMA*. Several unexpected facets of intellect also emerged as predictors of capacity to deal with the *high-demand* complexity shift (mean performance and slopes). Those with higher scores on the *Intellect-Seek* process and the

Intellect-Create domain were better equipped to deal with the *high-demand* complexity change, as well as three crossovers (*SeekThink*, *SeekCreate*, *ConquerCreate*).

4. Discussion

The current study aimed to develop a preliminary process account of CPS task performance by examining how non-cognitive factors influence performance trajectories on a microworlds CPS task above and beyond GMA. This aim was motivated by previous findings (e.g., Danner et al., 2011; Greiff et al., 2015; Greiff et al., 2012; Kretzschmar et al., 2016; Lotz et al., 2016; Sonnleitner et al., 2013; Wüstenberg et al., 2012), which have consistently found that CPS task performance explains variation in practical outcomes such as educational and workplace success beyond the variation explained by *g* or *Gf*. Our task, a CPS microworlds task, was designed to present participants with one block of trials segmented into three stages, each of increasing complexity to manage. Participants were not directly instructed how the task would change but instead needed to be sensitive to changes in the feedback interface as the task progressed – similar to how one would solve novel problems in a dynamic real-world environment. In line with the established relationships between intelligence, personality, and performance, we hypothesised that both GMA and specific non-cognitive capacities would have differential relationships with task performance as the task became more complex. We also hypothesised that certain intellectual investment and personality traits would influence mean performance and performance trajectories above and beyond GMA by motivating participants to engage with and deeply understand the microworld system. In summary, our findings support our broad hypothesis that CPS tasks capture both cognitive and non-cognitive capacities, however, our hypotheses on the changing role of specific abilities throughout the task were not all unambiguously supported.

Firstly, we found evidence for a complexity-GMA effect but not always in the expected direction. While we hypothesised that those higher in GMA would be less impacted by increases in task complexity, we found mixed evidence for this effect; high GMA conferred performance advantages initially but this was attenuated over time such that GMA did not confer performance benefits in the final and most challenging stage. GMA is a strong predictor of performance in high-pressure cognitive situations but a weaker predictor of performance in typical cognitive situations (von Stumm et al., 2011). Our CPS microworlds task was designed to reflect a more typical cognitive performance situation as it required participants to engage with and control a complex system over 61 trials; for most participants, this was intended to take around half an hour of continuous cognitive engagement. As it is difficult to sustain maximal performance for this period of time, it follows that this may be one reason as to why GMA became less predictive of performance over time.

Conversely, we found evidence to support our hypotheses that non-cognitive capacities would predict performance in the CPS microworlds task above and beyond a general mental ability factor. Most of the intellectual investment and personality traits we measured accounted for mean performance differences between stages when controlling for GMA, suggesting a broad role for non-cognitive factors in driving CPS microworlds performance. Further, examining how these traits influenced within-stage performance trajectories allowed us to understand how these traits facilitated learning and mastery of the task as it became more complex over time. Fewer factors facilitated this capacity to learn and master complexity increases on our microworlds task than were associated with mean performance differences.

We hypothesised that several personality and investment trait domains would aid both mean performance, and learning and mastery of rule changes—NFC, Learning Goal Orientation, Intellect, Open-Mindedness, and Conscientiousness—and that Negative Emotionality would hinder learning and mastery of rule changes. Our hypotheses were supported for Open-Mindedness and Conscientiousness, which conferred mean performance benefits for the *medium-* and *high-demand* effects, and partially supported for NFC, Learning Goal Orientation, and Intellect, which conferred mean performance benefits for the *high-demand* effect only. Our hypotheses were also supported for mean performance impairments for those high in Negative Emotionality for both effects. When we examined the role of these traits on influencing performance trajectories, the only hypotheses supported were for Intellect (*high-demand* effect) and Conscientiousness (*medium-demand* effect). Nonetheless, both Extraversion (*medium-demand* effect) and Performance-Prove Goal Orientation (*high-demand* effect) emerged as unexpected yet significant factors that aided learning and mastery of rule changes. At the trait facet level, we hypothesised that the Intellect-Conquer process, the Intellect-Learning and Intellect-Thinking operations, Open-Mindedness: Intellectual Curiosity and Creative Imagination, and Conscientiousness: Productiveness would also facilitate rule learning as complexity increased, and that the Emotional Volatility and Anxiety domains of Negative Emotionality would hinder rule learning as complexity increased. Of our facet-level hypotheses, only that for Conscientiousness: Productiveness (*high-demand* effect) was supported. Nevertheless, a number of other facets emerged as significant predictors of performance, specifically Conscientiousness: Organisation and Conscientiousness: Responsibility (*medium-demand* effects), Extraversion: Sociability and Extraversion: Energy Level (*medium-demand* effects), Extraversion: Sociability and Extraversion: Assertiveness Level (*high-demand* effects), Intellect Seek and Create (*high-demand* effects).

Interestingly, our findings suggest two distinct clusters of non-cognitive capacities that influence the ability to manage dynamic complexity changes in our microworlds task. These clusters stem from their role in either the *medium-demand* effect or the *high-demand* effect. The *medium-demand* effect represents the capacity to manage the shift from fixed to variable outflow of species from the island. Both mean performance and the strength and direction of performance trajectories differed between the low and medium complexity stages depending on an individual's level of Conscientiousness and its Organisation and Responsibility facets (both in the same direction). Specifically, a crossover effect in the performance trajectories was observed such that Conscientiousness (and Organisation and Responsibility) did not differentiate between performance trajectories in the first and easiest stage, however, those with low Conscientiousness displayed more rapid increases in penalty scores (i.e., slower rule learning and mastery) with the move from low to medium complexity than those high in Conscientiousness. This was illustrated in Figure 8. Both mean performance and the strength and direction of performance trajectories also differed between the low and medium complexity stages depending on individuals' level of Extraversion and its Energy Levels and Sociability facets; with both facet effects in the same direction but of differing strengths. For Extraversion as a trait and for both facets, there were no differences in the slope of performance trajectories by Extraversion level in the low complexity stage, but crossover effects were observed in the medium complexity stage. Those of lower Extraversion (and facets) displayed sharper increases in penalty score (i.e., slower rule learning and mastery) than those of higher Extraversion (and facets) with the move from low to medium complexity. This was illustrated in Figure 9.

Taken together, this suggests that those with high Conscientiousness and high Extraversion tended to manage the medium-demand complexity increase better than their low trait counterparts. This may stem from the possibility that high levels of both traits facilitate cognitive exploration. In this task, the only rule that can be precisely deduced is the fixed outflow rules in the first stage. In the second stage, shifting to variable outflow, participants are unlikely to deduce the specific nature of the rule (i.e. random outflow between 10 and 30 per trial) and even if they did, the random nature of species loss from the ecosystem would mean it would be impossible to perform "perfectly" (i.e. get a penalty score of 0 on each trial). Nevertheless, controlled cognitive exploration may facilitate the acquisition of an optimal control strategy. For example, keeping input to around 20 species per trial should theoretically control the system relatively well. This kind of controlled

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 July 2022 doi:10.20944/preprints202207.0246.v1

cognitive exploration is characteristic of those high in Conscientiousness—who are organised and systematic, and have a tendency toward reliable and steady behaviours (Soto & John, 2017)—and of those high in Extraversion—who tend to display more exploratory behaviours on challenging cognitive tasks (Greiff et al., 2019).

A different cluster of traits was associated with the capacity to manage the *high-demand* shift on our CPS microworlds task. These traits thematically represent a shift from controlled cognitive exploration to more creative exploration: NFC, Learning Goal Orientation, and Intellect. Furthermore, they also capture the capacity and tendency to perform in cognitive situations that can be emotionally taxing, indicated by the roles of Perform-Prove Goal Orientation and Neuroticism: Emotional Volatility and Anxiety. The shift to high complexity introduces decision delays to the microworld system, that is, a lag between decisions being made and taking effect that requires participants to consider how their actions have impacted the current state and will impact upon future trials. Combined with variable outflow, this again means it is not practically possible to perform perfectly. There is an added cognitive load of firstly needing to recognise there is a delay, secondly needing to quantify the extent of the delay, and thirdly needing to synthesise the information provided to make appropriately informed decisions. High creative exploration traits may facilitate more rapid acquisition of the presence of a delay by encouraging behaviours of seeking and trying new control strategies, and examining and implementing the feedback provided in the system interface (i.e., distance from goal) more rapidly. Indeed, this is reflected in our results, where NFC, Learning Goal Orientation, and Intellect were associated with *high-demand* mean performance differences. Furthermore, Intellect and both the Seek process and Create domain were associated with *high-demand* performance trajectory effects such those higher in these traits had shallower penalty score trajectories (i.e., they learned to control the system more rapidly). As noted in the results, we interpret the Intellect facet results with caution, however, the trait Intellect result remains interpretable.

Moreover, the shift from no delay to delay is likely to create uncertainty and confusion. This increased demand on cognitive resources may create stress on participants meaning that an individual's susceptibility to stress and capacity to manage it may become a determinant of performance in this final and challenging stage of the task. This added load may create performance anxiety, thus the buffering effect of low Emotional Volatility and Anxiety, as well as Performance-Prove Goal Orientation – which is consistently related to Neuroticism (McCabe et al., 2013). This is supported by our findings, in which Neuroticism facets Emotional Volatility and Anxiety were both not associated with performance trajectory differences for the *medium-demand* shift were associated with performance trajectory differences for the *high-demand* shift; penalty scores for those high in Emotional Volatility and Anxiety accumulated at a faster rate than those low in these trait facets suggesting these trait facets impeded rapid rule learning and mastery. This may be compounded by the fact that implicit feedback is available to participants in the form of their penalty score on the display, so they are readily able to see how they are performing on the task. As a result, those high in Performance-Prove Goal Orientation who are motivated by outperforming others to prove their own ability (VandeWalle, 1997), may be more motivated to discover the new delay rule and implement an effective control strategy to manage it. In addition, those high in Emotional Volatility and Anxiety are more sensitive to negative feedback (Hirsh & Inzlicht, 2008), further affecting their performance and leading to a sharper increase in penalty score over time.

We pre-emptively make comment on the size of the effects we have uncovered here. What is notable from the graphs in Figures 4 to 7 is that the size of between-stage differences in the relationship between the non-cognitive trait and performance (Figures 4 and 6) is much larger than the size of the strength and direction of between-stage differences in performance trajectories as a function of non-cognitive traits (Figures 5 and 7). Our CPS microworlds task was around 30 minutes long and was designed to represent a typical performance task demanding extended engagement with a cognitive challenge (i.e., participants needed to engage consistently for the full 30 minutes to successfully control the system). This short burst of cognitive engagement is similar to a typical performance scenario, for example, completing a task at school, university, or work. Small effects, such as those found in our study, can have large ramifications in the real world (Götz et al., 2022). Small bursts of cognitive engagement accumulate over time to create a typical pattern of performance for an individual. Extrapolating our findings to a day, week, or even a month, could show that non-cognitive traits are an extremely strong predictor of typical performance in complex and challenging real-world environments.

Strengths, Limitations, and Future Directions

A key strength of our study was the use of GLMM, which allows us to understand how non-cognitive traits influence overall performance differences between stages but also how these traits dynamically influence performance trajectories within and between stages. This analysis has allowed us to draw conclusions based on more robust underlying analysis than, for example, comparing mean performance or correlations between stages. Furthermore, our use of a large participant pool means we had substantial range in microworlds task performance, as well as a large range of GMA and self-reported non-cognitive traits. This makes our research somewhat more generalisable than that which uses restricted samples (e.g. only university students or upper level business managers).

Future studies could extend upon our findings by examining how non-cognitive traits influence performance on other types of CPS tasks. For example, another common form of CPS task uses multiple inputs and outputs (Beckmann et al., 2017) and it would be prudent to examine whether our findings generalise to systems with a different structure of input and output variables. We also only examined the role of single traits due to the complexity of our modelling. Future studies could examine how non-cognitive capacities interactively affect performance using more complex analytical techniques to fully elucidate the role of different non-cognitive traits on CPS performance.

5. Conclusions

Our aim was to begin to develop a process-driven account of CPS task performance. Motivated by the desire to understand what factors influence CPS task performance above and beyond *g*, our findings suggest that CPS microworlds demand a wide variety of non-cognitive personality and investment traits. Our results also imply that distinct clusters of traits may differentially drive performance for different levels of task complexity – one cluster that facilitates the capacity to manage system randomness and one cluster that facilitates the capacity to manage system delays. These results go some way to explaining why CPS tasks explain variation in practical outcomes beyond *g*; these tasks appear to demand both cognitive ability and an individual's motivation to engage, explore, and voluntarily exert cognitive effort in the face of challenges. We have also demonstrated that the use of generalized linear mixed-effects models is an effective way to build and test process-driven accounts of the factors that influence how individuals monitor feedback and control performance in CPS contexts. By introducing consideration of lower, facet-level intellectual investment and personality traits, we have further contributed to the extant literature by generating novel insights into potential important mechanisms underlying successful management of CPS task environments. Our results complement the extant body of literature on the predictive utility of CPS tasks for educational and workplace attainment by beginning to develop a process account of CPS task performance; it appears that CPS tasks may be better predictors of academic and workplace success than traditional measures of *g* or *Gf* because they demand not just cognitive ability but also the motivation and willingness to engage with and deeply understand cognitive challenges.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data available on request due to restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin, 102*(1), 3-27. https://doi.org/10.1037/0033-2909.102.1.3

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*(3), 288-318. https://doi.org/10.1037/0096-3445.117.3.288

Ackerman, P. L. (1996). A Theory of Adult Intellectual Development: Process, Personality, Interests, and Knowledge. *Intelligence, 22*(2), 227-257. https://doi.org/10.1016/S0160-2896(96)90016-1

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Beckmann, J. F., Birney, D. P., & Goode, N. (2017). Beyond Psychometrics: The Difference between Difficult Problem Solving and Complex Problem Solving. *Front Psychol, 8*, 1739. https://doi.org/10.3389/fpsyg.2017.01739

Birney, D. P., Beckmann, J. F., & Beckmann, N. (2019). Within-individual variability of ability and learning trajectories in complex problems. In D. J. McFarland (Ed.), *General and Specific Mental Abilities* (pp. 253-283). Cambridge Scholars Publishing.

Birney, D. P., Beckmann, J. F., Beckmann, N., Double, K. S., & Whittingham, K. (2018). Moderators of learning and performance trajectories in microworld simulations: Too soon to give up on intellect!? *Intelligence, 68*, 128-140. https://doi.org/10.1016/j.intell.2018.03.008

Brunner, J., & Austin, P. C. (2009). Inflation of Type I error rates in multiple regression when independent variables are measured with error. *The Canadian Journal of Statistics, 37*(1), 33-46.

Burns, B. D., & Vollmeyer, R. (2002, Jan). Goal specificity effects on hypothesis testing in problem solving. *Q J Exp Psychol A, 55*(1), 241-261. https://doi.org/10.1080/02724980143000262

Cacioppo, J., & Petty, R. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131. https://doi.org/10.1037/0022-3514.42.1.116

Condon, D. M., & Revelle, W. (2014). The International Cognitive Ability Resource: Development and initial validation of a public-domain measure. *Intelligence, 43*, 52-64. https://doi.org/doi.org/10.1016/j.intell.2014.01.004

Cripps, E., Wood, R. E., Beckmann, N., Lau, J., Beckmann, J. F., & Cripps, S. A. (2016). Bayesian Analysis of Individual Level Personality Dynamics. *Front Psychol, 7*, 1065. https://doi.org/10.3389/fpsyg.2016.01065

Danner, D., Hagemann, D., Schankin, A., Hager, M., & Funke, J. (2011). Beyond IQ: A latent state-trait analysis of general intelligence, dynamic decision making, and implicit learning. *Intelligence, 39*(5), 323-334. https://doi.org/10.1016/j.intell.2011.06.004

Dörner, D., & Funke, J. (2017). Complex Problem Solving: What It Is and What It Is Not. *Front Psychol, 8*, 1153. https://doi.org/10.3389/fpsyg.2017.01153

Fayn, K., Silvia, P. J., Dejonckheere, E., Verdonck, S., & Kuppens, P. (2019, Nov). Confused or curious? Openness/intellect predicts more positive interest-confusion relations. *J Pers Soc Psychol, 117*(5), 1016-1033. https://doi.org/10.1037/pspp0000257

Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning, 7*(1), 69-89. https://doi.org/10.1080/13546780042000046

Funke, J. (2010, May). Complex problem solving: a case for complex cognition? *Cogn Process, 11*(2), 133-142. https://doi.org/10.1007/s10339-009-0345-0

Gelman, A., Hill, J., & Yajima, M. (2012). Why we (usually) don't have to worry about multiple comparisons. *Journal of Research on Educational Effectiveness, 5*, 189-211.

Gonzalez, C., Vanyukov, P., & Martin, M. K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior, 21*(2), 273-286. https://doi.org/10.1016/j.chb.2004.02.014

Götz, F. M., Gosling, S. D., & Rentfrow, P. J. (2022). Small Effects: The Indispensable Foundation for a Cumulative Psychological Science. *Perspectives on Psychological Science, 17*(1). https://doi.org/10.1177/1745691620984483

Greiff, S., & Neubert, J. C. (2014). On the relation of complex problem solving, personality, fluid intelligence, and academic achievement. *Learning and Individual Differences, 36*, 37-48. https://doi.org/10.1016/j.lindif.2014.08.003

Greiff, S., Stadler, M., & Niepel, C. (2019). Extraversion, working style, reasoning, and complex problem solving: A study on the mechanisms underlying the link between extraversion and cognitive ability. *Psychological Test and Assessment Modeling, 61*(3), 321-332.

Greiff, S., Stadler, M., Sonnleitner, P., Wolff, C., & Martin, R. (2015). Sometimes less is more: Comparing the validity of complex problem solving measures. *Intelligence, 50*, 100-113. https://doi.org/10.1016/j.intell.2015.02.007

Greiff, S., Wüstenberg, S., & Funke, J. (2012). Dynamic Problem Solving. *Applied Psychological Measurement, 36*(3), 189-213. https://doi.org/10.1177/0146621612439620

Hirsh, J. B., & Inzlicht, M. (2008). The Devil You Know: Neuroticism Predicts Neural Response to Uncertainty. *Psychological Science, 19*(10). https://doi.org/10.1111/j.1467-9280.2008.02183.x

Hox, J. J. (2010). *Multilevel Analysis: Techniques and Applications* (2nd ed.). Routledge.

Hundertmark, J., Holt, D. V., Fischer, A., Said, N., & Fischer, H. (2015). System structure and cognitive ability as predictors of performance in dynamic system control tasks. *Journal of Dynamic Decision Making, 1*. https://doi.org/10.11588/jddm.2015.1.26416

IBMCorp. (2019). *IBM SPSS Statistics for Windows*. In (Version 26.0) IBM Corp.

Kretzschmar, A., Neubert, J. C., Wüstenberg, S., & Greiff, S. (2016). Construct validity of complex problem solving: A comprehensive view on different facets of intelligence and school grades. *Intelligence, 54*, 55-69. https://doi.org/10.1016/j.intell.2015.11.004

Lotz, C., Sparfeldt, J. R., & Greiff, S. (2016). Complex problem solving in educational contexts – Still something beyond a "good g "? *Intelligence, 59*, 127-138. https://doi.org/10.1016/j.intell.2016.09.001

McCabe, K. O., Van Yperen, N. W., Elliot, A. J., & Verbraak, M. (2013). Big Five personality profiles of context-specific achievement goals. *Journal of Research in Personality, 47*(6), 698-707. https://doi.org/10.1016/j.jrp.2013.06.003

Mussel, P. (2013, May). Intellect: a theoretical framework for personality traits related to intellectual achievements. *J Pers Soc Psychol, 104*(5), 885-906. https://doi.org/10.1037/a0031918

Rigas, G., & Brehmer, B. (1999). Mental Processes in Intelligence Tests and Dynamic Decision Making Tasks. In P. Juslin, Montgomery, H. (Ed.), *Judgment and Decision Making: Neo-brunswikian and Process-tracing Approaches*. Lawrence Erlbaum Associates.

Rudolph, J., Greiff, S., Strobel, A., & Preckel, F. (2018). Understanding the link between need for cognition and complex problem solving. *Contemporary Educational Psychology, 55*, 53-62. https://doi.org/10.1016/j.cedpsych.2018.08.001

Sonnleitner, P., Keller, U., Martin, R., & Brunner, M. (2013). Students' complex problem-solving abilities: Their structure and relations to reasoning ability and educational success. *Intelligence, 41*(5), 289-305. https://doi.org/10.1016/j.intell.2013.05.002

Soto, C. J., & John, O. P. (2017, Jul). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *J Pers Soc Psychol, 113*(1), 117-143. https://doi.org/10.1037/pspp0000096

Stadler, M., Becker, N., Gödker, M., Leutner, D., & Greiff, S. (2015). Complex problem solving and intelligence: A meta-analysis. *Intelligence, 53*, 92-101. https://doi.org/10.1016/j.intell.2015.09.005

Stadler, M., Niepel, C., & Greiff, S. (2019). Differentiating between static and complex problems: A theoretical framework and its empirical validation. *Intelligence, 72*, 1-12. https://doi.org/10.1016/j.intell.2018.11.003

R Core Team. (2019). *R: A language and environment for statistical computing*. In (Version 3.6.2) R Foundation for Statistical Computing. https://www.R-project.org/

VandeWalle, D. (1997). Development and Validation of a Work Domain Goal Orientation Instrument. *Educational and Psychological Measurement, 57*(6), 995-1015. https://doi.org/10.1177/0013164497057006009

Vollmeyer, R., & Rheinberg, F. (1999). Motivation and metacognition when learning a complex system. *European Journal of Psychology of Education, 14*(4), 541-554. https://doi.org/10.1007/BF03172978

von Stumm, S., & Ackerman, P. L. (2013, Jul). Investment and intellect: a review and meta-analysis. *Psychol Bull, 139*(4), 841-869. https://doi.org/10.1037/a0030746

von Stumm, S., Chamorro-Premuzic, T., & Ackerman, P. L. (2011). Re-Visiting Intelligence – Personality Associations: Vindicating Intellectual Investment. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell Handbook of Individual Differences* (1 ed.). Blackwell Publishing Ltd.

Wood, R. E., Clarke, T., Beckmann, J. F., & Birney, D. P. (2009). Simulations, learning and real world capabilities. *Education + Training, 51*(5/6), 491-510. https://doi.org/10.1108/00400910910987273

Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving — More than reasoning? *Intelligence, 40*(1), 1-14. https://doi.org/10.1016/j.intell.2011.11.003

Young, S. R., & Keith, T. Z. (2020). An Examination of the Convergent Validity of the ICAR16 and WAIS-IV. *Journal of Psychoeducational Assessment, 38*(8), 1052-1059. https://doi.org/10.1177/0734282920943455