THE WORK HIGH OF KNOWING WHY

An exploration into extrinsically and intrinsically framed task significance interventions and their effect on the emotional component of effortful tasks

Kate Thompson, MSc
Department of Psychological and Behavioural Science
The London School of Economics and Political Science
ABSTRACT

Whilst it is well known that a stronger sense of purpose in life can help improve our physical and mental health, we know surprisingly little about the wellbeing impacts of purpose in work. Beyond the economic benefits established in existing literature, does understanding the “why” behind our work help improve the emotional component of everyday jobs? To investigate this question, an online randomised controlled trial was conducted on 330 adults in the UK to test the impact of task significance interventions on the positive and negative affect induced by a real-effort task. Furthermore, to advance our understanding of optimum intervention formats, the between-subject design used two treatment arms to examine the variance between an extrinsic task significance frame, where significance is communicated to participants, and an intrinsic task significance frame, where significance is made by participants.

Statistical analysis indicates an empirical link between task significance interventions and increased positive affect following effortful work, and surprisingly suggests that an extrinsic purpose frame is superior to an intrinsic frame (positive affect=+3.152, SE=1.284, OLS p-value=0.015, Romano Wolf p-value=0.086, d=3.273). In response to these findings, a new model, the Trinity of Task Significance Sources, is offered to help inform future investigations looking to optimise task significance interventions. This study cautiously contributes to the gaps in workplace purpose literature surrounding affect and the associated theoretical frameworks of eudemonic wellbeing in work. Furthermore, in the context of the workplace mental-health crisis facing the UK today, the suggestion that cost-effective, scalable task significance interventions could help individuals feel 12% increased positive emotions in work (shifting them from below to above population average positive affect) presents several policy and organisational implications stretching across work-related environments from offices to classrooms.

Key Words: Eudemonic Wellbeing, Work Purpose, Task Significance, Positive Affect, Negative Affect, Workplace Mental-Health
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RCT</td>
<td>Randomised Controlled Trial</td>
</tr>
<tr>
<td>TS</td>
<td>Task Significance</td>
</tr>
<tr>
<td>PA</td>
<td>Positive Affect</td>
</tr>
<tr>
<td>NA</td>
<td>Negative Affect</td>
</tr>
<tr>
<td>n</td>
<td>Total number of individuals or observations in the sample size</td>
</tr>
<tr>
<td>H_a</td>
<td>Alternative Hypothesis</td>
</tr>
<tr>
<td>H_0</td>
<td>Null Hypothesis</td>
</tr>
<tr>
<td>DV</td>
<td>Dependant Variable</td>
</tr>
<tr>
<td>IV</td>
<td>Independent Variable</td>
</tr>
<tr>
<td>FWER</td>
<td>Family Wise Error Rate</td>
</tr>
<tr>
<td>rwolfp</td>
<td>Romano Wolf Adjusted P-Value correcting for multiple hypotheses</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>SE</td>
<td>Standard Error</td>
</tr>
<tr>
<td>R^2</td>
<td>R-Squared</td>
</tr>
<tr>
<td>R^2_adj.</td>
<td>Adjusted R-Squared</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS

Abstract .......................................................................................................................... 2

List of Abbreviations .................................................................................................. 3

Table of Contents ......................................................................................................... 4

Table of Figures & Tables ............................................................................................ 6

1.0 Introduction ............................................................................................................ 7

2.0 Literature Review .................................................................................................. 10

   2.1 Conceptualising Purpose .................................................................................. 10

   2.2 Purposeful Work .............................................................................................. 13

   2.3 Task Significance and Affect ........................................................................... 16

   2.4 Extrinsic and Intrinsic Task Significance & the Act of Meaning-Making .......... 18

3.0 Hypotheses ............................................................................................................ 20

4.0 Methods ................................................................................................................ 21

   4.1 Research Subjects & Sample Size .................................................................. 21

   4.2 Experimental Design ...................................................................................... 22

   4.3 Platform & Distribution .................................................................................. 27

   4.4 Reward & Incentives ...................................................................................... 28

   4.5 Framework for Data Analysis ......................................................................... 28

5.0 Results .................................................................................................................. 30

   5.1 Sample Responses & Characteristics .............................................................. 30

   5.2 Hypothesis Testing ........................................................................................... 31

6.0 Discussion .............................................................................................................. 42

7.0 Conclusion ............................................................................................................. 48

8.0 References ............................................................................................................. 50

9.0 Appendices ............................................................................................................ 62

   9.2 Ethical Guidelines & Data Safeguarding ........................................................... 63

   9.3 Sample Size Justification and Realisation ......................................................... 64
9.4 Multiple Hypotheses Correction Testing ........................................... 65
9.5 Exploratory Analyses ........................................................................... 67
9.6 Survey Detail ......................................................................................... 68
TABLE OF FIGURES & TABLES

**Figure 4.2.1:** Summary of Procedure Flow .......................................................... 23
**Figure 4.2.3:** Participant View of Slider-Sorting Task on Qualtrics ..................... 26
**Figure 5.1:** Sample Characteristics Summary Statistics ........................................... 32
**Figure 5.2.1a:** Histograms Visualising PA by Condition ........................................ 32
**Figure 5.2.1b:** Boxplot Showing PA Spread and Centres by Condition .................. 32
**Figure 5.2.1c:** Histograms Visualising NA by Condition ......................................... 32
**Figure 5.2.1d:** Boxplot Showing NA Spread and Centres by Condition .................. 33
**Table 5.2.1a:** Multiple Linear Regression Results for All Task Significance Treatment
Effects on Positive Affect ......................................................................................... 35
**Table 5.2.1b:** Multiple Linear Regression Results for All Task Significance Treatment
Effects on Negative Affect ......................................................................................... 35
**Figure 5.2.2a:** Histograms Visualising PA by Condition ........................................ 36
**Figure 5.2.2b:** Boxplot Showing PA Spread and Centres by Condition .................. 36
**Figure 5.2.2c:** Histograms Visualising NA by Condition ......................................... 37
**Figure 5.2.2d:** Boxplot Showing NA Spread and Centres by Condition .................. 37
**Table 5.2.2a:** Multiple Linear Regression Results for Extrinsic and Intrinsic Task
Significance Treatment Effects on Positive Affect .................................................... 39
**Table 5.2.2b:** Multiple Linear Regression Results for Extrinsic and Intrinsic Task
Significance Treatment Effects on Negative Affect .................................................... 39
**Figure 6.2:** Trinity of Task Significance Sources ..................................................... 47
**Table 9.4.1:** Romano Wolf Multiple Hypothesis Corrected P-Values For Simple Linear
Regressions ................................................................................................................. 66
**Table 9.4.2:** Romano Wolf Multiple Hypothesis Corrected P-Values For Multiple Linear
Regressions ................................................................................................................. 66
**Figure 9.6:** Original Slider Task Schematic .............................................................. 68
1.0 INTRODUCTION

The influence of purpose on human behaviour is an enduring concept dating back to Aristotle’s theories of eudaimonia (Aristotle, 350 B.C.E/1925), yet one we still don’t fully understand. The concept of purposeful work has emerged as a prominent construct for institutions and organisations linked to positive wellbeing and economic outcomes such as increased performance (Allan, 2017), job satisfaction (Steger et al., 2012), engagement, and reduced levels of burn out (Fairlie, 2011). Scholars have therefore looked to identify antecedents of purposeful work to inform behavioural interventions (Rosso et al., 2010). From this research, task significance (TS) – increasing the extent to which employees believe their work benefits the world beyond them (Kulik & Oldham, 1987) - has consistently emerged as an effective behavioural intervention, with studies establishing a causal link between increased TS and increased productivity (Ariely et al., 2008), helping behaviours (Grant, 2008) and higher quality work outcomes (Chandler & Kapelner, 2013) across workforces (Salamone & Lordan, 2022).

There are, however, two gaps within the TS literature that merit exploration. Firstly, whilst studies have explored the relationship between overall sense of life purpose and emotional reactions to everyday stressors (Hill et al., 2018), we understand very little about the influence a sense of work purpose has on the emotional component of performing everyday tasks. The first objective of this research is therefore to examine the role work purpose plays in improving affect experienced during effortful activity, with the aim of empirically establishing a scalable, cost-effective intervention that improves overall affect at work, enabling the auxiliary benefits that better workplace affect is a determinate of, such as enhanced executive function (Lerner et al., 2004; Mittal & Ross, 1998), ethical decision-making (Loewenstein & Lerner, 2003; Weber & Johnson, 2009) and increased employee psychological and physiological wellbeing (Hooker et al., 2018; Sin et al., 2015).
Secondly, we know very little about the most effective format for a TS intervention, with existing research having only examined TS narratives that are extrinsically framed, involving an external body communicating a pre-determined purpose to the individual. This contrasts with the innate human will to seek meaning for ourselves (Frankl, 1946), and the literature substantiating the benefits of intrinsic motivation over extrinsic motivation (Ryan & Deci, 2000). Therefore, with a second research objective to advance our understanding of optimum TS interventions, this study compares the effect of an extrinsic TS frame (where significance is communicated to participants) and an intrinsic TS frame (where significance is made by participants) on affect and productivity associated with effortful tasks.

To examine these two objectives, a randomised controlled trial (RCT) was conducted to test the impact of increasing the salience of task significance on the positive and negative affect induced by a real-effort task. A between-subject design involved 3 arms, one control and two treatment conditions, to compare extrinsic and intrinsic TS frames. Statistical analysis indicates an empirical link between task significance interventions and increased positive affect following effortful work, and surprisingly suggests that an extrinsic purpose frame is superior to an intrinsic frame (positive affect=+3.152, SE=1.284, OLS p-value=0.015, Romano Wolf p-value=0.086, d=3.273). In response to these findings, a new model, the Trinity of Task Significance Sources, is offered to help inform future investigations looking to optimise task significance interventions.

This investigation is particularly timely when set against the backdrop of the growing workplace wellbeing epidemic facing the UK. With Stevenson’s (2017) governmental review suggesting that 40% of the UK workforce experience work-related mental health issues annually, generating an estimated yearly economic loss of £34.9bn to employers and £24 – £27bn to government bodies (Deloitte, 2017), it is clear that more must be done to identify cost-effective, scalable interventions to help tackle the workplace wellbeing crisis, with the potential to stretch across multiple work-related applications from offices to classrooms. Alongside these organisational and policy implications, this
research aims to contribute to the literature on: (1) organisational behavioural research focused on purpose and its effect on workplace wellbeing (Martela & Pessi, 2018; Rosso et al., 2010b); (2) theoretical frameworks of eudemonic wellbeing, particularly the optimum ingredients required to generate task significance (R. Baumeister & Vohs, 2002; Ryff, 1989; Waterman, 1990); (3) compensating wage theory specifically the role of task meaning differentials (Ariely et al., 2008; Rosen, 1987).

This paper is structured as follows: Chapter 2 reviews the related literature, setting out existing research gaps and the empirical context for the study. Associated hypotheses for testing are summarised in Chapter 3, with Chapter 4 overviewing the RCT methodology, including research subjects, sampling, experimental design, and framework for data analysis. Chapter 5 presents hypothesis testing results, with Chapter 6 discussing possible explanations, limitations, and recommendations for future research. Chapter 7 concludes by summarising key findings.
2.0 LITERATURE REVIEW

2.1 Conceptualising Purpose

Whilst a lack of consistent definitions within the literature has led to complexity and experimental limitations (Dolan & White, 2007), most scholars begin by examining purpose in the context of meaning. Baumeister’s (1991) book devoted to meaning in life builds on foundational philosophical works from Aristotle to Locke to Herbar, offering a definition that benefits from breadth and depth: meaning is a “shared mental representation of possible relationships among things, events, and relationships. Thus, meaning connects things” (p. 15). Baumeister later further dimensionalised this: “meaning can link two things even if they are physically separate entities” (Baumeister & Vohs, 2002, p. 628). This uniquely human ability to connect two seemingly separate things has led to intangible but socially crucial structures, from the use of money to the social justice system.

It is important to recognise that meaning can exist across multiple levels, and, as examined in Vallacher & Wegner’s (1987) Action Identification Theory, our behaviour is shifted when these levels alter. Vallacher & Wegner define the interacting principles that identify a causal interdependence between what we think we are doing and what we do, categorising 3 different levels of meaning: low levels focus on immediate, empirical, and often precise facts (i.e., physically driving my car to the grocery shop), whereas high levels of meaning can transcend across time and connect current experiences to far-off goals (i.e., shopping to nourish my growing family and raise healthy children). Through this framework we can start to recognise the relationship between meaning and purpose.

To support that relationship, we must also turn to Baumeister & Wilson’s (1996) model of the Four Needs of Meaning which shape the narrative structures we create for ourselves to help make sense of life (Mcadams, 2001). Whilst needs number two,
three and four are needs for values, a sense of efficacy and self-worth, unsurprisingly, the first is the need for purpose. They observe that whilst we colloquially default to a single purpose life-narrative, i.e., ‘the meaning of life’, practically, most of us hold tension between several purposes spanning relationships, family, work, faiths, and meaningful projects. Whilst less coherent, this multiplicity buffers us from potential meaninglessness if one domain is impaired, such as job loss or divorce. Ultimately these multiple purposes serve as “themes around which the life story can be constructed. That is, specific events can be understood in relation to the purpose, and their meaning consists of how they either promote or hamper movement towards that purpose” (Baumeister & Wilson, 1996, p. 2).

However, whilst the concepts of purpose and meaning are inextricably connected, it is impossible to conceptualize purpose without examining its relationship to happiness. In Bradburn’s (1969) early distinction between positive and negative affect, he critically chose to label the outcome variable “happiness”, validating that choice by referencing Aristotle’s (1985) Nicomachean Ethics, which attests the highest human achievement as eudaimonia. This semantic link led to a consequential operationalising of the word “happiness” as a shorthand for the mental state account of affect (Parfitt, 1984), which, alongside a cognitive evaluation of life, combines to form our modern definition of subjective wellbeing (Diener, 1999; Kahneman & Krueger, 2006). However, as many have since distinguished, a truer translation of the Greek word eudaimonia is not simply happiness, but the notion that a fulfilled life only comes from activities that allow us to realise the potential of our true selves, our “daimon” (Dolan, 2015).

This experience of eudaimonia is characterised by self-expression (A. S. Waterman, 1990), by realising our unique potential as individuals and a species (A. Waterman, 1993), and by a belief that one’s life and activities have direction (Ryff, 2013; Ryff & Singer, 2008), meaning, (Frankl, 1946) and worthwhileness (Kashdan et al., 2008). Whilst they are not opposites, a happy life (“overall, how happy did you feel
yesterday?”) and purposeful life (“overall, to what extent do you feel that the things you do in your life are worthwhile?”) are clearly not the same (Dolan et al., 2008; John F. Helliwell, 2017).

Whilst, compared to hedonic wellbeing, this eudemonic facet of wellbeing remains surprisingly neglected within empirical literature, a small but growing body of research has begun to explore the positive outcomes and behaviours resulting from increased sense of life purpose, linking it to physical and mental health benefits such as reductions in depression and Alzheimer’s risk (Boyle, Buchman, Barnes, et al., 2010; Wood & Joseph, 2010), incident disability (Boyle, Buchman, & Bennett, 2010), incidents of stroke (Kim et al., 2013) and even improved life-expectancy (Boyle et al., 2009). Studies have also linked life purpose with positive increases in cognitive function, such as improved determination (Hill et al., 2016), self-esteem (Burrow & Rainone, 2017), identify formation (Bronk, 2011) and consideration of society beyond ourselves (McKnight & Kashdan, 2009).

When questioning why purpose has the influence established in these studies, interesting foundations can be found in Klinger’s (1998) work on Evolutionary Goal Theory, which explores our instinctive need to seek out what is required for survival. Klinger suggests that because this “persistent goal striving, which constitutes an imperative of purpose” has characterised our history, “within evolutionary theory, every feature of animal organisms must have evolved in the service of goal pursuit” (Klinger, 1998, p. 27). Whilst it is important to resist tying up psychological explanations in a neat evolutionary bow, it appears that creating meaning from ostensibly disconnected stimuli is neurologically wired into our brains. In studying split-brain surgery patients, Gazzaniga et al. (1996) noticed that patients whose fibers (which normally join the two brain hemispheres) had been separated were no longer able to transmit information between hemispheres, but instead used their left brains to interpret their received information with narrative reflection, which he termed "left-brain interpreter" (Baumeister & Vohs, 2002). Furthermore, Rice’s (2019) research
exploring goal-achieving experiences in rat brains observed dopamine spikes in the nucleus accumbent and ventromedial prefrontal cortex (associated, in human brains, with subjective value of rewards during decision-making), not only when a goal was achieved but also when it was set. Unsurprisingly, the larger the goal achievement, the larger the biochemical spike. Whether the explanations are anthropological or neurological, it is clear we are a species wired to seek purpose.

2.2 Purposeful Work

Amongst the drivers of purpose, evidence demonstrates that work is particularly critical for perceived purpose in life, but is also an unoptimized ingredient (Steger et al., 2012). In comparison to past centuries of pastural or domestic work, where tasks were identical and goal-realisation immediate (i.e., food on the table), the 20th century’s industrial revolution reinvented working landscapes, introducing varied careers with often intangible goals, sometimes several psychological steps removed from the ultimate organisational outcome (Carton, 2018). In response to this, the study and systemisation of purposeful work has been a defining theme for modern employers and policy makers. For the purposes of this research, the concept of purposeful work is again understood in its eudemonic (rather than hedonic) terms with a positive valance towards meaningfulness (Steger et al., 2012).

Whilst many theoretical frameworks have been offered to explain why more purposeful work fundamentally impacts our behaviour (Cameron, 2012; Haslam et al., 2000; Martela & Pessi, 2018; Pratt et al., 2003; Wrzesniewski et al., 2003), Rosso et al’s (2010) model of the core psychological and social mechanisms through which work becomes meaningful offers several strengths. Not only is it widely referenced as an explanatory source within empirical investigations (i.e., Salamone & Lordan, 2022), with several studies proving Rosso’s suggested mechanistic relationships (Allan et al., 2018a; Grant, 2008), it also provides a unique integration of decades of defining related theories, from Jung’s (1997) formulation of the individual and Erikson’s (1994)
psychosocial stages, to Rogers’ (1961) study of the fully functioning person, Maslow’s (2020) construct of self-actualisations and even Ryff’s (1989) structure for positive psychological functioning. Rosso’s seven mechanisms for purposeful work are: 1) authenticity: a coherence between our behaviour and our real selves, 2) self-efficacy: our belief that we can make a difference, 3) self-esteem: the assessment of our self-worth, 4) purpose: connection to our direction in life, 5) belongingness: our desire for acceptance in a community, 6) transcendence: connecting to something beyond our selves and 7) cultural sense-making: the sociocultural forces that shape our interpretation of meaning.

Alongside efforts to define the frameworks explaining why purposeful work influences how we behave, there is growing evidence linking work purpose to economic and welfare benefits. Workers with more purpose contribute more discretionary unpaid hours (Wrzesniewski et al., 1997), energy and effort (Bunderson & Thompson, 2009) and commitment (Duffy et al., 2014), whilst also accepting lower wages (Steger et al., 2010). Additionally, research demonstrates increased manager-rated and self-rated performance (Allan et al., 2018a), better colleague cohesion (Sparks & Schenk, 2001) and reduced levels of burn out (Fairlie, 2011). Increased work meaning is also linked to greater job satisfaction (Kamdron, 2005) and engagement (Kahn, 1990). Surprisingly, in a multitopic General Social Survey, Cascio (2003) found that the belief that our daily work has meaning is more important to us than job features such as role security, promotion opportunities, working hours and even income.

With these compelling benefits in mind, scholars and practitioners have sought scalable and cost-effective interventions to increase everyday purpose in work. Hackman & Oldham’s (1976) Job Characteristics Model was one of the first, and still widely used, models to classify meaningfulness in job design, defining skill variety, task identity and task significance as antecedents of meaningful jobs. Of these, task significance (TS) – increasing an individual’s perception that their task has purpose because it positively improves the welfare of others within and outside the
organisation (Allan et al., 2018b) - has led the way in behavioural intervention design, with several meta-analyses finding robust correlations between the specific variable of TS and increased perception of meaningful work (Allan, 2017).

Whilst detailed understanding of how TS impacts behaviour and wellbeing is nascent, some studies have explored several positive outcomes effected by TS. Particularly noteworthy for this paper is Ariely et al.’s (2008) Lego experiments which first investigated TS on labour supply, establishing the causal link between TS, increased performance and lower wage demands in a lab setting. Participants (n=20) in the Meaningful condition (where Lego models were kept) built 3.4 more models with $0.39 lower reservation wage, compared to participants (n=20) in the Sisyphus condition (where Lego was dismantled) (SD unavailable, p-value=0.005).

Grant’s (2008) experiments with university fundraising callers and lifeguards corroborated this productivity effect in the field, and established causal relationships between TS, increased job dedication and helping behaviours (results reviewed in Appendix 9.4 as part of sample calculations). Critically, his work is the first to identify that, beyond perceived meaningfulness (Allan, Duffy, et al., 2018a), two additional relational mechanisms that mediate the effects of TS are increased perceptions of social worth and social impact – thus reinforcing the role of self-esteem, self-efficacy and transcendence outlined in Rosso’s model above.

Chandler & Kapelner’s (2013) pioneering studies must also be noted as the first to replicate these findings in a scaled natural field setting online, recruiting 2471 workers via Amazon MTurk for data-entry. They found workers in the meaningful group (who were told data-entry was needed to study tumoral cells) were 8.8% more likely to label +5 images (SE:2.4, p=0.024) compared to the control condition (whose work was shredded). Salamone & Lordan’s (2022) work must also be highlighted for establishing these effects in a blue collar setting with factory workers. However, whilst these studies have moved a once non-existent field forward, there remain significant gaps in
our understanding across work purpose more broadly, but particularly within the narrower field of TS. For example, whilst links to economic benefits (such as productivity and job dedication) have been established, we have yet to examine wellbeing benefits that TS interventions could impact.

2.3 Task Significance and Affect

Whilst research examining links between work purpose and affect is uncharted, studies have explored the relationship between broader life purpose and affect (Hicks et al., 2012; King et al., 2006; Thoresen et al., 2003), with several relevant findings. For example, Schaefer et al’s (2013) work links purpose in life to emotional resilience in the form of faster recovery from negative stimulus, and Zilioli et al’s (2015) study reveals purpose in life predicts decreased rates of allostatic load (a biomarker that measures dysregulation across multiple physiological systems over time). However, limitations exist within these studies as they rely on measuring the purpose in life variable via national happiness data, which is not only subjective but reveals correlated rather than causal relationships, e.g., Hill et al (2018) who correlate sense of purpose to stress reduction using a subset of the US Midlife reports.

Whilst the purpose in work variable is easier to control, measure and influence, to this researcher’s knowledge there have been no empirical investigations researching the impact of purposeful work, specifically TS, on affect experienced in work. Burrow et al (2016) must be noted for their study establishing how a purposeful treatment reduced mountain climbers’ sense of steepness and perceived physical effort, however, whilst ‘work-like’, this sits outside the employment domain. We must also note Allan, Dexter, et al’s (2018) study on the impact of meaningful work on mental health components of depression, anxiety, and stress, however, affect is not specifically explored, nor is a work-like simulation (e.g., real data-entry) used.
Which begs the question, why is it important to examine the impact of TS on workplace affect? The answer is found in the extensive literature establishing a link between increased positive affect (PA) and compelling auxiliary benefits including physiological health (Vazquez, 2017; Wood & Joseph, 2010) such as reduced inflammation (Sin et al, 2015), lower chronic condition rates (Hooker et al., 2018), physical resilience (Smith et al., 2009) and improved wellbeing (Burton & King, 2009; Lyubomirsky et al., 2005). Literature also indicates PA’s impact on cognitive function such as increased working memory (Figueira et al., 2018), increased quality of decision making (Bechara & Damasio, 2005; Loewenstein & Lerner, 2003; Peters et al., 2006), reduced risk taking (Mittal & Ross, 1998) and more ethical decisions (Weber & Johnson, 2009), which have an important baring on workforce performance. Finally, in the context of work specifically, PA has been linked to various benefits including increased work satisfaction (Billings & Isen, 1989), lower turnover intentions and higher employee influence (Mittal & Ross, 1998; Park et al., 2021).

The findings in these studies can be explained through several theoretical frameworks that examine interactions between affect and wellbeing at work, such as Weiss & Cropanzano’s (1996) Affective Events Theory which examines how workers’ internal influences (i.e., mental states, emotions, cognitions) connect to their reactions to events in external work environments, ultimately impacting job performance and satisfaction. Together, these frameworks plus the growing research linking emotions to wellbeing and economic impacts, provide a robust rationale for further exploration into scalable workplace interventions that optimise affect. Furthermore, considering the research suggesting that experienced meaningfulness might only partly mediate the link between TS and performance (Humphrey et al., 2007), there is merit in exploring the role of affect as an additional mechanism through which TS might travel.
2.4 Extrinsic and Intrinsic Task Significance & the Act of Meaning-Making

The further gap in TS literature exists in relation to its format as an intervention. In most studies, TS treatments adopt similar formats with common ingredients of 1) social worth / recognition, 2) task application and 3) social impact / transcendence to participants, normally via a narrative text to communicate that information (Ariely et al., 2008; Chandler & Kapelner, 2013; Grant, 2008; Salamone & Lordan, 2022). Which raises the question as to whether alternative intervention formats exist and how we can deliberately assess what the optimum format is. In particular, the literature has yet to examine different TS frames, for example exploring whether there is variance between a frame where significance is communicated to the individual, versus a frame where significance is made by the individual. As humans, we are not only wired to seek meaning from others but, critically, we are wired to make it for ourselves. This specific act of meaning-making can be defined as the active search for significance (Park & Folkman, 1997), and as a process through which people evaluate or re-evaluate an event or series of events (Taylor, 1983).

Positive outcomes have been linked to meaning-making interventions that prompt people to undergo subjective sense-making or benefit-finding activities. These have been established in the clinical world as faster recovery from traumatic events (Pennebaker et al., 1988), reduced physician visits (Pennebaker & Beall, 1986), improved bereavement processing (Currier et al., 2006) and cancer-diagnosis processing (Chan & Ho, 2007) and even improved HIV recoveries (Bower et al., 1998). In work, meaning-making interventions have been linked to better academic results (Pennebaker et al., 1990), improved in-role performance and adaptability in the workplace (Van den Heuvel et al., 2020).

Ryan & Deci’s (1985) Self Determination Theory helps to explain why meaning-making interventions have this influence, examining the social-contextual conditions that fuel or foil our natural processes of self-motivation, and identifying three innate human
needs of competence, autonomy, and relatedness. Interestingly, their theory also links these factors to enhancement in intrinsic motivation, which we can draw on as an adjacent but closely related body of psychological literature that supports the hypothesis that subjective TS made might deliver more impact than objective TS received. Building on dual-process models (Kahneman, 2013), Ryan & Deci (2000) distinguish between extrinsic motives, which require controlled information processing, and intrinsic motives which concern innate, often automatic processing. We can borrow from these classifications to define a new distinction between an extrinsic TS frame, where an individual is communicated a controlled, objective purpose of the task and how it positively impacts others (the prime format used in the literature), and an alternative intrinsic TS frame, that prompts an individual to make their own innate, subjective meaning, connecting the task with their purpose and reflecting on how they believe they are positively impacting others through it.
3.0 HYPOTHESES

This study examines two research questions: (1) Does increasing the salience of task significance increase the positive affect or reduce the negative affect an effortful task induces? (2) Does framing task significance intrinsically improve affect or productivity during effortful tasks more than framing it extrinsically?

Hypothesis 1: Effect of Task Significance Interventions on Affect

- $H_a1$ Compared to the Control group, the Treatment Groups will reveal increased positive affect or reduced negative affect resulting from effortful tasks

Hypotheses 2: Effect of Extrinsic vs Intrinsic Task Significance Frame on Affect

- $H_a2a$ Compared to the Control group, the Extrinsic Task Significance group will reveal increased positive affect or reduced negative affect resulting from effortful tasks
- $H_a2b$ Compared to the Control group, the Intrinsic Task Significance group will reveal increased positive affect or reduced negative affect resulting from effortful tasks

Hypotheses 3: Effect of Extrinsic vs Intrinsic Task Significance Frame on Productivity

- $H_a3a$ Compared to the Control group, the Extrinsic Task Significance group will reveal increased productivity resulting from effortful tasks
- $H_a3b$ Compared to the Control group, the Intrinsic Task Significance group will reveal increased productivity resulting from effortful tasks

Hypothesis 4: Affect as a Mechanism

- $H_a4$ Data analysis reveals affect as a hidden mechanism through which task significance travels

---

1 Note, whilst the content and number of hypotheses tests remain, following Pre-Registration, a revision request was made to my supervisor to tighten up the hypothesis groupings, which was approved.
4.0 METHODS

To test these hypotheses, an online randomised controlled trial (RCT) with a between-subject design was conducted, in which participants randomly received either an extrinsically framed or intrinsically framed task significance treatment before completing the real-effort task of sorting through computer sliders. The emotional component of real-effort was then measured via PANAS Affect Questionnaire (Watson et al., 1988) and by revealing participants’ willingness to accept (WTA) payment to repeat the task. Productivity was measured in the form of task scores.

4.1 Research Subjects & Sample Size

Sample size calculations were computed to mitigate against risk of Type I and II errors whilst ensuring the RCT had enough statistical power to reveal the minimum detectable effect. Using results from Grant’s (2008) comparable study on the impact of TS, using stories to communicate task social benefits to callers soliciting alumni contributions to their university, G-Power was used for sample size computations, revealing Total n=315 (Condition 0=105, Condition 1=105, Condition 2=105). See Appendix 9.3 for detailed sample calculations and justification.

This random sample of 315 UK participants were recruited from Prolific Academica, chosen for its validated data quality, measured in comprehension, attention, and honesty, compared to alternative platforms (Peer et al., 2022). The sample was disturbed to evenly balance male and female participants to enable future gender-based sub-group exploratory analysis following gender differences established in life purpose (Xi et al., 2022). Several pre-screening criteria were administered. To ensure compliance with LSE’s Research and Ethics Policy and Procedures, participants were pre-screened as over 18, and, before commencing, were provided with the research objectives, task and time requirements, reward structure and asked explicitly to provide Informed Consent (Appendix 9.6.2). Participants who had engaged in past pilot
iterations of this study were excluded. To minimise the additional distractions surrounding mobile phone usage, participants were limited to desktop or laptop participation only. To ensure comprehension of the nuanced framing within treatments, participants were pre-screened for English as a first language. Finally, to ensure high quality participation, individuals were pre-screened for Prolific approval rating >95 and history of previous submissions >500.

4.2 Experimental Design

4.2.1 Procedure Overview

The experiment was designed in survey format, using Qualtrics software. 3 rounds of 1 week-long pilot iteration sprints were conducted (n=45: n=15 in each pilot) to iterate the procedure for optimum participant experience and experiment efficacy. As visualised in Figure 4.2.1 below, survey structure incorporated a 4-part procedure:

1) **Set-Up**: informed consent and attention check
2) **Treatment Conditions**: random assignment into: (i) Control Condition (ii) Extrinsic Task Significant Condition or (iii) Intrinsic Task Significance Condition
3) **Real-Effort Task**: 6 rounds x 48 slider-sorting
4) **Affect Measurement & Demographic Capture**:
   I. Affect measurement: (i) PANAS Positive and Negative Affect Subscales (Watson & Clark, 1988) (ii) Willingness to Accept revelation
   II. Demographic questions: age, gender, ethnicity, education, income, employment, and seriousness checks
4.2.2 Treatment Conditions

Extrinsic and Intrinsic TS treatments were both introduced in the form of additional contextual information plus reflection questions, designed to (1) convey meaningful TS (2) avoid deception whilst also not revealing the actual treatment (3) avoid experimenter demand. Following the design of Chandler & Kapelner (2013), these were embedded into the introduction section of the real-effort task.

The Extrinsic TS treatment content introduced additional background on the research aims and outcomes, framed from the researcher’s perspective “we wanted to thank you again for taking part today, and explain why your contribution of time and effort in this experiment makes a real difference”, with two subsections explaining the researchers’ purpose for this experiment and the impact they hope it will achieve. The content covered key ingredients necessary to motivate significance, outlined in Section 2.3, (i) Social Worth / Recognition: e.g., “thank you” “your contribution of time and commitment”, (ii) Task Application: e.g., “contribute to the literature around mental health” and (iii) Social Impact / Transcendence: e.g., “help the growing epidemic of workplace wellbeing issues”. Following this content, participants had 5 minutes to
answer 3 compulsory open-ended questions: 1) “*What is our motivation for asking you to take part in this scientific research today?*” 2) “*Please take a moment to think about our purpose and goals in this study. How does your participation in the research help contribute to these?*” 3) “*What impact can this research have on others?*”.

The Intrinsic TS treatment content introduced a shorter research goal summary, without the detail found in the Extrinsic treatment, to avoid crowding out intrinsic TS. Instead, the text focused on prompting participants to make their own meaning: “*we thought it was important to give you a moment to reflect on why you are personally taking part in this scientific study*”. The same compulsory timed 3 open-ended questions supported this, with almost identical wording to the Extrinsic treatment, other than replacing ‘our/we’ with ‘your/you’: (1) “*What is your motivation for taking part in this scientific research today?*” (2) “*Please take a moment to think about your personal purpose and goals in life. How does an activity like participating in this research help you contribute to these?*” (3) “*What impact can you have on others by taking part in this research?*”.

The Control treatment omitted additional research goal context but controlled for the variable of a disclosive writing experience, by focusing the same 3 compulsory timed open-ended questions on the unrelated topic of the last movie seen by participants, following the design of Burrow & Hill (2013). See Appendix 9.6.2 for full treatments.

In all conditions, the use of open-ended reflective questions augments Burrow & Hill’s (2013) experiment, built on evidence that engaging in disclosive writing about self-regulatory topics (such as purpose in life) is associated with wellbeing benefits (King, 2001; Park, 2010), and rooted in Pennebaker’s (1993) explanation that during disclosive writing, the use of insight and causal words associated with cognitive activity brings a simplified, more coherent cognitive structure to remembered or anticipated experiences.
4.2.3 Real-Effort Task

Following treatments, all participants were then directed to the same real-effort task designed to induce the exertion of genuine effort to simulate a real-world work scenario. To achieve this, a slider-sorting task was chosen for its validation in the literature (Bosman et al., 2005; Choi et al., 2018). Compared to alternative real-effort tasks, the significant merits of slider-sorting are (1) it is easy and quick to explain, which is critical outside of a lab setting, (2) it requires no prior knowledge thus avoiding confounding results with variability of skill levels (3) therefore it supports more broadly generalisable results across labour markets, and (4) it remains identical across rounds, involving minimal randomness (Gill & Prowse, 2012).

The specific task design was taken from Gill & Prowse’s (2011) work where participants were asked to sort sliders into a central mark of 50 on a horizontal scale on lab computers (Appendix 9.6.1). In their work, participants completed a total of 6 rounds with 48 sliders in each round, timed at 120 seconds per round. Participants were told their total points score at the end would be the total number of successfully sorted sliders across all rounds. In this RCT, piloting revealed that the slider functionality allowed participants to override the instructions by clicking on a central point (i.e., 50) for a faster, exact drop, which risked limiting the ability to control the real-effort variable across the group. Therefore, the task was modified to fit Qualtrics software code with an evolved design requiring participants to adjust sliders to exact numbers set in that slider’s label (see Figure 4.2.3 below).
A countdown timer was also introduced, increasing pressure to induce additional real-effort. A limitation of this feature was the potential for both time pressure plus real-effort to increase attrition, therefore the pilots played a critical role in carefully calibrating the optimal amount of pressure to task feasibility. The addition of clear, simple instructions and a practice task round was also carefully iterated during pilot phase to increase participants’ confidence and familiarity before the first round. See Appendix 9.6.2 for full task details.

4.2.4 Affect Measurement

Primary affect measurement was captured through the PANAS Affect Questionnaire (Watson et al., 1988), chosen for its wide applications, well-established construct validity and high correlations with equivalent measures of combined state affect (Crawford & Henry, 2004). Immediately following the task, participants completed 20
questions reflecting on momentary experience of combined Positive and Negative Affect Subscales: Interested; Distressed; Excited; Upset; Strong; Guilty; Scared; Hostile; Enthusiastic; Proud; Irritable; Alert; Ashamed; Inspired; Nervous; Determined; Attentive; Jittery; Active; Afraid. Answers were captured using 5-point Likert scale: 1) Very Slightly or Not at All 2) A Little 3) Moderately 4) Quite a Bit 5) Extremely.

A secondary proxy affect measurement was introduced in the form of willingness to accept (WTA), where participants were asked how much money they would accept to complete the task again, ranging from £0.00 - £20.00.

4.3 Platform & Distribution

An online setting was chosen due to its ability to achieve the sample size required in short timeframes and modest budget constraints. Compared to lab environments, several additional benefits of online include access to a larger, more diverse sample and reduced time investment burden of participants. Moreover, participating from a home computer more closely simulates the flexible home-working environment many experience today. However, internet platforms also raise unique limitations. Alongside general online issues of increased attrition, repeat participation and attention issues (which have been addressed in the above design and sample selection methodology), the lack of control over the experimental environment (Skitka & Sargis, 2006), presents a unique challenge to affect measurement in this study. Three measures were therefore introduced to partially address this. Firstly, participants were not allowed to complete the experiment on mobiles or tablets, to reduce notification distractions and environmental changes that come from device use in transit. Secondly, participants were asked to complete the study in one sitting and were screened out if they left and returned later. Thirdly, the exploratory measure of WTA was introduced as a proxy measure of affect to mitigate against these environmental limitations.
Prolific’s data safeguarding procedure was upheld, alongside a thorough Data Management Plan adhering to the LSE and Government’s Data Ethics Framework (Appendix 9.2).

4.4 **Reward & Incentives**

Rewards were utilised in accordance with Prolific’s ethical compensation recommendations which enforce minimum hourly rewards of £6.00 / $8.00, depending on the effort required. The study did not involve a niche target population, nor required outside-platform effort, however it did require participation in a real-effort task. Therefore, an increase to £7.03/hr (£3.05 for estimated completion time of 26 minutes) was offered as an incentive to foster required levels of engagement and provide high data quality. An additional bonus incentive was deliberately excluded to ensure the isolated effect of TS was not impaired by introducing conflicting economic motivations.

4.5 **Framework for Data Analysis**

4.5.1 **Dependent Variables**

There were two dependent variables (DV) used in primary analysis and one in exploratory analysis. The primary measure of affect was Positive Affect (PA) “papost” and Negative Affect (NA) “napost”, calculated using the momentary PANAS scoring-system justified above. Total PA and NA Subscale Scores ranged from 10 – 50, where 10 is the lowest i.e., all 10 subscale questions answered “very slightly or not at all”. Higher PA scores reflect higher PA; higher NA scores reflect higher NA. For reference, average population momentary PA score is 29.7 (SD=7.9), whilst average population momentary NA is lower at 14.8 (SD=5.4) (Boyle, 2015). It must be noted that whilst PANAS captures full-spectrum affect, guidelines require independent analysis of the PA/NA Subscales, as less NA is not equivalent to more PA (Watson & Clark, 1988).
An exploratory proxy measure of affect “wta” was captured as participants’ willingness to pay to repeat the task, in £0.50 increments ranging from £0.00 to £20.00. The primary measure of productivity, “scores”, summed the total number of sliders that participants correctly sorted within 6 rounds of the task, with a total possible range of 0 – 288.

4.5.2 Independent and Control Variables

Independent variables corresponded to the 3 conditions: (i) Control Condition, “treat==0”; (ii) Extrinsic TS Condition, “treat==1”; (iii) Intrinsic TS Condition, “treat==1”. To compare combined all TS treatments against Control, dummy variables “allts==1” and “allts==0” were created. Additional control variables that may contribute to the variability of affect and productivity were also measured in standard socio-economic demographics of “gender”, “age”, “ethnicity”, “education”, and workplace demographics of “income” and “employment”.

4.5.3 Statistical Analysis

Statistical analysis was performed using Stata/SE 17.0, see Do File in Appendix 9.7. Hypothesis testing was conducted using simple linear regression analysis alongside paired t-tests, with multiple linear regressions to control for effect of covariates on outcomes of interest. Normality was checked using the Shapiro Wilk test, and results were validated with robustness checks including outlier analysis, and performance of the Mann Whitney test and the Kolmogorov-Smirnov test.

Significance was assessed at the 0.05 level. The Romano Wolf correction was administered for 9x multiple hypotheses, to control the Family Wise Error Rate (FWER) and allow for dependence among p-values by bootstrap resampling (Clarke et al., 2019). Where statistical significance was found in OLS model p-values, Romano Wolf adjusted p-values (rwolfp) have been reported alongside them. See Appendix 9.4 for more detail.
5.0 RESULTS

5.1 Sample Responses & Characteristics

A total of 336 responses were collected, with all participants agreeing to Informed Consent and passing attention and seriousness checks. Participants using a mobile phone or tablet (n=6) were excluded as survey functionality required desktop or laptop devices. Of the 330 total, Control n=112, Extrinsic Task Significance =105, Intrinsic Task Significance n=113. Within this sample, age ranged from 19-82 with mean 46.9, there was 50/48.79 balance between male/female, and ethnicity broadly reflected the UK’s national composition. 66.06% of the sample were in full time or part time work, with 14.55% retired. See detail in sample characteristics summary statistics in Table 5.1.

Table 5.1: Sample Characteristics Summary Statistics

<table>
<thead>
<tr>
<th>Age</th>
<th>Mean</th>
<th>SD</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 18-24</td>
<td>3.724</td>
<td>1.395</td>
<td>16</td>
<td>4.85</td>
</tr>
<tr>
<td>2. 25-34</td>
<td>57</td>
<td>17.27</td>
<td>73</td>
<td>22.12</td>
</tr>
<tr>
<td>3. 35-44</td>
<td>78</td>
<td>23.64</td>
<td>68</td>
<td>20.61</td>
</tr>
<tr>
<td>4. 45-54</td>
<td>38</td>
<td>11.52</td>
<td>38</td>
<td>11.52</td>
</tr>
<tr>
<td>5. 55-64</td>
<td>38</td>
<td>11.52</td>
<td>38</td>
<td>11.52</td>
</tr>
<tr>
<td>6. 65+</td>
<td>38</td>
<td>11.52</td>
<td>38</td>
<td>11.52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>Mean</th>
<th>SD</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Male</td>
<td>1.521</td>
<td>0.563</td>
<td>165</td>
<td>50</td>
</tr>
<tr>
<td>2. Female</td>
<td>161</td>
<td>48.79</td>
<td>161</td>
<td>48.79</td>
</tr>
<tr>
<td>3. Non binary/3rd</td>
<td>2</td>
<td>6.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Self describe</td>
<td>1</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Prefer not to say</td>
<td>1</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Mean</th>
<th>SD</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. White/caucasian</td>
<td>1.461</td>
<td>1.496</td>
<td>291</td>
<td>88.18</td>
</tr>
<tr>
<td>2. Asian-Eastern</td>
<td>6</td>
<td>1.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Asian-Indian</td>
<td>11</td>
<td>3.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. African-American</td>
<td>10</td>
<td>3.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Mixed Race</td>
<td>4</td>
<td>1.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Other</td>
<td>4</td>
<td>1.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Prefer not to say</td>
<td>4</td>
<td>1.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Education</th>
<th>Mean</th>
<th>SD</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Masters or above</td>
<td>2.327</td>
<td>0.811</td>
<td>150</td>
<td>45.00</td>
</tr>
<tr>
<td>2. Bachelors</td>
<td>142</td>
<td>43.03</td>
<td>142</td>
<td>42.64</td>
</tr>
<tr>
<td>3. Secondary</td>
<td>120</td>
<td>36.26</td>
<td>120</td>
<td>36.26</td>
</tr>
<tr>
<td>4. Other</td>
<td>16</td>
<td>4.85</td>
<td>16</td>
<td>4.85</td>
</tr>
<tr>
<td>5. Prefer not to say</td>
<td>2</td>
<td>0.61</td>
<td>2</td>
<td>0.61</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Income</th>
<th>Mean</th>
<th>SD</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. £25k</td>
<td>2.215</td>
<td>1.223</td>
<td>96</td>
<td>29.09</td>
</tr>
<tr>
<td>2. £25 - £50k</td>
<td>131</td>
<td>39.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. £50 - £100k</td>
<td>78</td>
<td>23.64</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. £100 - £200k</td>
<td>5</td>
<td>1.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. £200k+</td>
<td>1</td>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Prefer not to say</td>
<td>19</td>
<td>5.76</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Employment</th>
<th>Mean</th>
<th>SD</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Working full-time</td>
<td>2.739</td>
<td>2.194</td>
<td>154</td>
<td>46.67</td>
</tr>
<tr>
<td>2. Working part-time</td>
<td>64</td>
<td>19.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Unemployed</td>
<td>11</td>
<td>3.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Homemaker</td>
<td>18</td>
<td>5.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Student</td>
<td>10</td>
<td>3.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Retired</td>
<td>40</td>
<td>14.55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Other</td>
<td>21</td>
<td>6.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Prefer not to say</td>
<td>4</td>
<td>1.21</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.2 Hypothesis Testing

5.2.1 Hypothesis 1: Effect of Task Significance Interventions on Affect

\( H_0 \) Compared to the Control group, the Treatment Groups will reveal increased positive affect or reduced negative affect resulting from effortful tasks

The total (n=330) mean PA is 29.197 (standard deviation [SD]=9.659), with minimal skew (0.110) and mild kurtosis (2.185). Median PA is 29, with a range of 10 – 50. The total (n=330) mean NA is 12.094 (SD=4.389), with a right skew (4.296) and very high kurtosis (27.295), suggesting non-normal distribution. Median NA is 10, with a range of 10 – 50, suggesting several outliers (<2xSDs from mean). These right skewed, non-normal findings are unsurprising, bearing in mind population momentary average is close to 10 at only 14.8 (SD=5.4), suggesting most people experience average low state NA (Boyle, 2015). All TS treatment group (n=218) has a higher mean PA of 29.995 (SD=9.890) compared to control group (n=112) PA mean 27.643 (SD=9.034) and a higher mean NA of 12.431 (SD=5.056) compared to control mean NA of 11.438 (SD=2.539).\(^2\) See summary statistics visualised in Histograms (\textit{Figure 5.2.1a, 5.2.1c}) and Boxplots showing spread and centres (\textit{Figure 5.2.1b, 5.2.1d}) below.

\(^2\) Whilst unequal sample sizes of control group (n=112) vs combined all-treatments group (n=218) must be noted, ANOVA tests were used for additional robustness checks.
Figure 5.2.1a: Histograms Visualising PA by Condition

Histograms display total PA PANAS scores following real effort task within Control and All Task Significance Treatment Conditions across a range of 0 – 50. Navy line over imposes group means. Blue line over impose a normal distribution.

Figure 5.2.1b: Boxplot Showing PA Spread and Centres by Condition

Boxplot showing spread and centres of PA PANAS data set across Control and All Task Significance Treatment Conditions.

Figure 5.2.1c: Histograms Visualising NA by Condition

Histograms display total NA PANAS scores following real effort task within Control and All Task Significance Treatment Conditions across a range of 0 – 50. Navy line over imposes group means. Blue line over impose a normal distribution.
Assuming normality, a simple linear regression model reveals that PA increases by 2.353 (Standard Error[SE]=1.117) for participants who received TS interventions with statistical significance (p-value=0.000), suggesting we can reject H0:1 of no TS treatment effect on PA. A two-tailed t-test supports these findings (p-value=0.036, t-ratio=2.105) and results are upheld in ANOVA tables. Robustness checks were conducted using multiple linear regression with inclusion of covariates “age”, “income”, “ethnicity”, “employment” (sample was balanced for gender), selected on pairwise correlations (p-value<0.05) and impact on Adjusted R-Squared [R²adj]. Compared to the univariate model, inclusion of these exogenous variables still reveals a higher PA of 2.272 (SE=1.099) for participants receiving the intervention, with retained significance (p-value=0.039) and increased exploratory power (R-Squared [R²]=6.10%), suggesting that we can reject H0:1 that there is no TS treatment effect on PA. Significance was upheld during outlier analysis. However, it is important to note that when correcting for 9x multiple hypotheses, significance does not withstand (rwolfp= 0.176) and thus we must handle these results with caution. Shapiro-Wilk tests (p-value=0.00) suggest we can reject the null that PA data is normally distributed, however, tests indicate that the residuals are normally distributed (see Figure 5.2.1e), suggesting OLS as an appropriate analytical model. The Mann Whitney Test was still run for additional robustness checks, revealing significance (p-value=0.039), validating these OLS findings.
A simple linear regression model reveals that NA increases by 0.993 (SE=0.508) for participants who received TS interventions, however, without significance (p-value=0.051). A two-tailed t-test supports these findings (t-ratio=1.956, p-value=0.051) and results are upheld in ANOVA tables. Compared to the univariate model, multiple linear regression with covariates “age”, “income”, “ethnicity”, “employment” reveals 1.070 (SE=0.504) higher NA for participants receiving treatment. Whilst, compared to the simple linear model, effects are now significant (p-value=0.035), with increased explanatory power ($R^2=4.06\%$), we must apply caution bearing in mind non-normal NA data distribution displayed in Figure 5.2.1c. Shapiro-Wilk tests p-value=0.00 further suggest a rejection of $H_0$ that data is normally distributed, and further tests indicate residuals are not normally distributed (Figure 5.2.1e), confirming OLS as inappropriate. The Mann Whitney Test reveals no significance (p-value=0.654), suggesting that we cannot reject $H_0$1 that there is no TS treatment effect on NA. A Kolmogorov-Smirnov test, performed to test equality of distributions, and an Ordered Probit Model, performed as a Likert appropriate regression to test maximum likelihood, were conducted for additional robustness, and uphold these findings (all p-values>0.050).

See Table 5.2.1a and 5.2.1b for $H_0$1 Multiple Linear Regression Results
Table 5.2.1a: Multiple Linear Regression Results for All Task Significance Treatment

Effects on Positive Affect

<table>
<thead>
<tr>
<th>Positive Affect</th>
<th>Coef.</th>
<th>St.Err.</th>
<th>t-value</th>
<th>p-value</th>
<th>[95% Conf Interval]</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>All TS Treatments</td>
<td>2.272</td>
<td>1.099</td>
<td>2.07</td>
<td>.039</td>
<td>.111</td>
<td>4.433</td>
</tr>
<tr>
<td>age</td>
<td>1.52</td>
<td>.397</td>
<td>3.83</td>
<td>0</td>
<td>.74</td>
<td>2.301</td>
</tr>
<tr>
<td>income</td>
<td>.151</td>
<td>.43</td>
<td>0.35</td>
<td>.726</td>
<td>-.695</td>
<td>.996</td>
</tr>
<tr>
<td>ethnicity</td>
<td>.553</td>
<td>.358</td>
<td>1.55</td>
<td>.123</td>
<td>-.151</td>
<td>1.258</td>
</tr>
<tr>
<td>employment</td>
<td>-.011</td>
<td>.247</td>
<td>-0.04</td>
<td>.966</td>
<td>-.497</td>
<td>.475</td>
</tr>
<tr>
<td>Constant</td>
<td>20.922</td>
<td>2.06</td>
<td>10.15</td>
<td>0</td>
<td>16.869</td>
<td>24.975</td>
</tr>
</tbody>
</table>

Mean Positive Affect (PA) | 29.197 | SD Positive Affect (PA) | 9.659 |
R-squared | 0.061 | Number of obs | 330 |
F-test | 4.209 | Prob > F | 0.001 |
Akaike crit. (AIC) | 2423.536 | Bayesian crit. (BIC) | 2446.330 |

*** p<.01, ** p<.05, * p<.1. Multiple regression means are adjusted for covariates of age, income, employment, and ethnicity. See above for Romano Wolf multiple hypothesis p-value corrections.

Table 5.2.1b: Multiple Linear Regression Results for All Task Significance Treatment

Effects on Negative Affect

<table>
<thead>
<tr>
<th>Negative Affect</th>
<th>Coef.</th>
<th>St.Err.</th>
<th>t-value</th>
<th>p-value</th>
<th>[95% Conf Interval]</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>All TS Treatments</td>
<td>1.071</td>
<td>.504</td>
<td>2.12</td>
<td>.035</td>
<td>.078</td>
<td>2.063</td>
</tr>
<tr>
<td>age</td>
<td>-.361</td>
<td>.182</td>
<td>-1.98</td>
<td>.048</td>
<td>-.719</td>
<td>-.003</td>
</tr>
<tr>
<td>income</td>
<td>-.328</td>
<td>.197</td>
<td>-1.66</td>
<td>.098</td>
<td>-.716</td>
<td>.061</td>
</tr>
<tr>
<td>ethnicity</td>
<td>.106</td>
<td>.164</td>
<td>0.64</td>
<td>.52</td>
<td>-.218</td>
<td>.429</td>
</tr>
<tr>
<td>employment</td>
<td>.177</td>
<td>.113</td>
<td>1.56</td>
<td>.119</td>
<td>-.046</td>
<td>.4</td>
</tr>
<tr>
<td>Constant</td>
<td>12.817</td>
<td>.946</td>
<td>13.55</td>
<td>0</td>
<td>10.956</td>
<td>14.679</td>
</tr>
</tbody>
</table>

Mean Negative Affect (NA) | 12.094 | SD Negative Affect (NA) | 4.388 |
R-squared | 0.041 | Number of obs | 330 |
F-test | 2.744 | Prob > F | 0.019 |
Akaike crit. (AIC) | 1909.937 | Bayesian crit. (BIC) | 1932.732 |

*** p<.01, ** p<.05, * p<.1. Multiple regression means are adjusted for covariates of age, income, employment, and ethnicity. Negative Affect data-set follows non-normal distribution, with Mann-Whitney Tests revealing insignificant findings. See above for Romano Wolf multiple hypothesis p-value corrections.

5.2.2 Hypothesis 2: Effect of Intrinsic vs Extrinsic Task Significance Frame on Affect

- H₀₂a Compared to the Control group, the Extrinsic Task Significance group will reveal increased positive affect or reduced negative affect resulting from effortful tasks

- H₀₂b Compared to the Control group, the Intrinsic Task Significance group will reveal increased positive affect or reduced negative affect resulting from effortful tasks
Extrinsic TS treatment (n=105) has higher mean PA, 30.962 (SD=10.163) compared to control (n=112), 27.643 (SD=9.034), and a higher mean NA, 11.819 (SD=3.257), compared to control, 11.438 (SD=2.539). Intrinsic TS treatment (n=113) has higher mean PA, 29.097 (SD=9.587) compared to control, 27.643 (SD=9.034), and a higher mean NA, 13.000 (SD=6.245), compared to control, 11.438 (SD=2.539). See summary statistics visualised in Histograms (Figure 5.2.2a, 5.2.2c) and Boxplots showing spread and centres (Figure 5.2.2b, 5.2.2d) below.

**Figure 5.2.2a: Histograms Visualising PA by Condition**

![Histograms](image1)

Histograms display total PA PANAS scores following real effort task within Control, Extrinsic and Intrinsic Significance Treatment Conditions across a range of 0 – 50. Navy line over imposes group means. Blue line over impose a normal distribution.

**Figure 5.2.2b: Boxplot Showing PA Spread and Centres by Condition**

![Boxplots](image2)

Boxplot showing spread and centres of PA PANAS data set across Control, Extrinsic and Intrinsic Task Significance Treatment Conditions.
Assuming normality, a simple linear regression model reveals that PA increases by 3.319 (SE=1.303) for participants who received Extrinsic TS treatment, with statistical significance (p-value=0.011), suggesting we can reject \( H_0 \) that there is no Extrinsic TS treatment effect on PA. A two-tailed t-test supports these findings (t-ratio=-2.546, p-value=0.012) and results are upheld in ANOVA tables. Compared to the univariate model, inclusion of covariates “age”, “income”, “employment”, “ethnicity” (see rationale above) reduces the effect to 3.152 (SE=1.284), with retained significance (p-value=0.015) and increased exploratory power (\( R^2=6.6\% \)), suggesting that we can reject \( H_0 \) that there is no Extrinsic TS treatment effect on PA. Significance was upheld during outlier analysis. However, when correcting for 9x multiple hypotheses, findings are not significant at 0.05, but reveal indication of significance at 0.01 (rwolfp=0.086), thus we must handle these results with caution. As above, whilst...
Shapiro-Wilks tests suggest non-normal distribution (p-value=0.00) *Figure 5.2.1e* indicates the residuals are normally distributed, supporting OLS. Mann Whitney Tests were still run for additional robustness checks, revealing significance (p-value=0.016), validating OLS findings.

A simple linear regression model reveals that NA increases by 0.382 (SE=0.591) for participants who received Extrinsic TS interventions, with no statistical significance (p-value=0.519), suggesting that we cannot reject $H_{02a}$ that there is no Extrinsic TS treatment effect on NA. A two-tailed t-test support these findings (p-value=0.335, t-ratio=-0.966) and results are upheld in ANOVA tables and Mann Whitney Tests performed for additional robustness checks due to non-normal distribution of NA dataset established above (all p-values>0.05).

A simple linear regression model reveals that PA increases by 1.454 (SE=1.279) for participants who received Intrinsic TS interventions, with no statistical significance (p-value=0.256), suggesting that we cannot reject $H_{02b}$ that there is no Intrinsic TS treatment effect on PA. A two-tailed t-test confirms these findings (p-value=0.243, t-ratio=-1.171) and results are upheld in ANOVA tables and multiple linear regressions performed for additional robustness checks (all p-values>0.05). A simple linear regression model reveals that NA increased by 1.563 (SE= 0.580) for participants who received Intrinsic TS treatment, with statistical significance (p-value=0.007). These results are upheld by a two-tailed t-test (t-ratio=-2.454, p-value=0.015), suggesting we can reject $H_{02b}$ that there is no Intrinsic TS treatment effect on NA, however caution must be applied as, as established above, the NA dataset follows non-normal distribution (see *Figure 5.2.1e*). Mann Whitney Tests conducted for robustness, reveal no significance (p-value=0.654), suggesting that we cannot reject $H_{02b}$ that there is no Intrinsic TS treatment effect on DV NA. See *Table 5.2.2a and 5.2.2b* for $H_{02}$ Multiple Linear Regression Results.
Table 5.2.2a: Multiple Linear Regression Results for Extrinsic and Intrinsic Task
Significance Treatment Effects on Positive Affect

<table>
<thead>
<tr>
<th>Positive Affect</th>
<th>Coef.</th>
<th>St.Err.</th>
<th>t-value</th>
<th>p-value</th>
<th>[95% Conf Interval]</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extrinsic TS</td>
<td>3.152</td>
<td>1.284</td>
<td>2.45</td>
<td>.015</td>
<td>.625 - 5.678</td>
<td>**</td>
</tr>
<tr>
<td>Intrinsic TS</td>
<td>1.463</td>
<td>1.257</td>
<td>1.16</td>
<td>.245</td>
<td>-1.01 - 3.936</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>1.496</td>
<td>.397</td>
<td>3.77</td>
<td>.001</td>
<td>.716 - 2.277</td>
<td>***</td>
</tr>
<tr>
<td>income</td>
<td>.155</td>
<td>.429</td>
<td>.36</td>
<td>.718</td>
<td>-.689 - 1</td>
<td></td>
</tr>
<tr>
<td>ethnicity</td>
<td>.572</td>
<td>.358</td>
<td>1.60</td>
<td>.111</td>
<td>-.132 - 1.276</td>
<td></td>
</tr>
<tr>
<td>employment</td>
<td>-.01</td>
<td>.247</td>
<td>-0.04</td>
<td>.968</td>
<td>-.495 - .476</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>20.968</td>
<td>2.058</td>
<td>10.19</td>
<td>0</td>
<td>16.918 - 25.017</td>
<td>***</td>
</tr>
</tbody>
</table>

Mean Positive Affect (PA) 29.197, SD Positive Affect (PA) 9.659
R-squared 0.066, Number of obs 330
F-test 3.805, Prob > F 0.001
Akaike crit. (AIC) 2423.762, Bayesian crit. (BIC) 2450.356

*** p<.01, ** p<.05, * p<.1 Multiple regression means are adjusted for covariates of age, income, employment, and ethnicity. See above for Romano Wolf multiple hypothesis p-value corrections.

Table 5.2.2b: Multiple Linear Regression Results for Extrinsic and Intrinsic Task
Significance Treatment Effects on Negative Affect

<table>
<thead>
<tr>
<th>Negative Affect</th>
<th>Coef.</th>
<th>St.Err.</th>
<th>t-value</th>
<th>p-value</th>
<th>[95% Conf Interval]</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extrinsic TS</td>
<td>.489</td>
<td>.588</td>
<td>.83</td>
<td>.406</td>
<td>-.668 - 1.646</td>
<td></td>
</tr>
<tr>
<td>Intrinsic TS</td>
<td>1.606</td>
<td>.576</td>
<td>2.79</td>
<td>.006</td>
<td>.473 - 2.738</td>
<td>***</td>
</tr>
<tr>
<td>age</td>
<td>-.345</td>
<td>.182</td>
<td>-1.90</td>
<td>.058</td>
<td>-.703 - .012</td>
<td>*</td>
</tr>
<tr>
<td>income</td>
<td>-.331</td>
<td>.197</td>
<td>-1.68</td>
<td>.093</td>
<td>-.718 - .056</td>
<td></td>
</tr>
<tr>
<td>ethnicity</td>
<td>.093</td>
<td>.164</td>
<td>0.57</td>
<td>.569</td>
<td>-.229 - .416</td>
<td></td>
</tr>
<tr>
<td>employment</td>
<td>.177</td>
<td>.113</td>
<td>1.56</td>
<td>.119</td>
<td>-.045 - .399</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>12.787</td>
<td>.943</td>
<td>13.57</td>
<td>0</td>
<td>10.933 - 14.641</td>
<td>***</td>
</tr>
</tbody>
</table>

Mean Negative Affect (NA) 12.094, SD Negative Affect (PA) 4.388
R-squared 0.051, Number of obs 330
F-test 2.910, Prob > F 0.009
Akaike crit. (AIC) 1908.249, Bayesian crit. (BIC) 1934.842

*** p<.01, ** p<.05, * p<.1 Multiple regression means are adjusted for covariates of age, income, employment, and ethnicity. Negative Affect data set follows non-normal distribution, with Mann-Whitney Tests revealing insignificant findings. See above for Romano Wolf multiple hypothesis p-value corrections.
5.2.3 **Hypothesis 3: Effect of Intrinsic vs Extrinsic Task Significance Frame on Productivity**

- *Hₐ3a* Compared to the Control group, the Extrinsic Task Significance group will reveal increased productivity resulting from effortful tasks
- *Hₐ3b* Compared to the Control group, the Intrinsic Task Significance group will reveal increased productivity resulting from effortful tasks

Total RCT (n=330) mean productivity scores is 158.464 (SD=39.007), with mild right skew (0.252) and kurtosis (3.248). Median scores is 157, with range of 49 – 273. Intrinsic TS group (n=113) mean scores, 159.230 (SD=37.268), is higher than Extrinsic TS group (n=105) mean scores, 158.791 (SD=41.981), and both are higher than control group (n=112) mean, 158.383 (SD=38.153). See summary statistics visualised in Histograms (*Figure 5.2.3a*) and Boxplots showing spread and centres (*Figure 5.2.3b*) below.

*Figure 5.2.3a: Histograms Visualising Productivity Scores by Condition*

![Histograms](image)

Histograms display total productivity scores following real effort task within Control, Extrinsic and Intrinsic Significance Treatment Conditions across a range of 0 – 288. Navy line over imposes group means. Blue line over impose a normal distribution.
Figure 5.2.3b: Boxplot Showing Scores Spread and Centres by Condition

Two tailed t-tests reveal that, compared to control group mean (scores=27.643, SD=0.854), Extrinsic TS scored 1.407 (SE=5.440) higher mean and Intrinsic TS scored 1.856 (SE=5.028) higher mean, however neither were significant (p-values>0.05) suggesting that we cannot reject $H_{03a}$ or $H_{03b}$ that there is no Extrinsic or Intrinsic TS treatment effect on scores. ANOVA tables uphold these findings, as do simple and multiple linear regression analysis (with and without outliers) (all p-values>0.05).

5.2.3 Hypothesis 4: Affect as a Mechanism

$H_4$ Data analysis reveals affect as a hidden mechanism through which task significance travels

As results do not support $H_3$, relevant data required to perform a moderation of process analysis where $x=TS$ Treatment, $y=Productivity$ and $z=Affect$ are not complete. Thus we cannot reject $H_{04}$ that affect is a mechanism through which task significance travels.

See Appendix 9.5 for exploratory analysis examining the proxy affect DV, willingness to accept.
6.0 DISCUSSION

6.1 Does increasing the salience of task significance increase the positive affect or reduce the negative affect an effortful task induces?

This RCT’s goal was to study two research questions, the first of which examines the impact TS has on affect associated with effortful tasks. Overall, the findings supported H₂₁, in that increasing the salience of work purpose via TS interventions increases PA following effortful tasks (PA=2.27, SE=1.099, p-value=0.039, rwolfp=0.176, d=2.448). Whilst caution must be applied in acknowledging the loss of significance when correcting for 9x multiple hypothesis, these initial results suggest that literature linking purpose in life to increased PA (Schaefer et al., 2013) might also be applied to purpose in work, with the key distinction being that, unlike purpose in life (which is hard to influence as a variable), purpose in work can be impacted cost-effectively through TS interventions. Putting a real-world value on this potential effect, workers who understand the broader benefits of their task felt 8.5% increased PA following effortful work (emotions associated with pleasurable engagement with the environment such as “Inspired”, “Enthusiastic”, “Determined”, “Interested”) compared to those who don’t. To put that effect into further context, TS interventions have the potential to shift participants from below UK population momentary average PA to above average PA (UK Population PA average=29.7 [SD=7.9]; Control PA average=27.643 [SD=9.034], All TS PA average=29.995 [SD=9.890]) (J. Boyle, 2015). It is worth noting that whilst unprompted, even the control subjects may experience some sense of purpose (from participating in a scientific experiment), thus exhibiting lower NA or higher PA than in a field environment, therefore, for both groups the outcome is possibly downward biased.

Whilst promising, what is of particular interest is the one-sided nature of the affect impact. Whilst PA rose, NA (i.e., “Nervous”, “Jittery”, “Upset”, “Distressed”) did not fall, contradicting the literature linking purpose in life to reduced anxiety (Burrow et
One explanation might be that with a total participant mean momentary NA score, 12.094 (SD=4.389), below population momentary mean NA average score, 14.8 (SD=5.4) (Boyle, 2015) and close to PANAS baseline 10, scope for variance in NA reduction was limited overall. An alternative explanation could be found in the theories of Affect Balance (Bradburn 1969), Emotional Ambivalence (Fong & Tiedens, 2002) and Positivity Ratio (Fredrickson & Losada, 2005), where, whilst traditional research overwhelmingly considers PA and NA independently, these scholars make compelling cases to consider their interplay and coexistence in the context of wellbeing. These theories have been applied to a workplace context, with Yoon et al’s (2022) recent study examining the chemistry between PA and NA at work and associating it with a range of workplace wellbeing indices. This study found that at work, it is high PA, and not low NA, that drives life satisfaction, but that low NA at work is differentially linked to physical health. With this in mind, perhaps optimum affect balance, rather than simply NA reduction, is the desirable affect outcome from TS interventions? Future experiments could benefit from a more forensic examination of this interaction between PA and NA following TS interventions.

It is important, however, to apply caution to these findings by recognising several limitations within this study. The multiple hypotheses tested here (exacerbated by the independence of PA and NA Subscales) pose a critical statistical weakness which must be addressed with future research examining the singular hypotheses to categorically establish significant intervention effects. The fundamental design flaw of self-reported measurement must also be noted. It poses the issue of inconsistency and unreliability, if participants cannot accurately assess themselves (Pfister & Böhm, 2008), challenging the internal validity of the design. Whilst much literature exploring correlations between affect and life purpose are built on the same shaky foundations, and whilst PANAS has been thoroughly validated as an affect measure, it is critical that future research replicates these findings in a lab using objective biological affect measures such as facial recognition, HRV (heart rate variability) or EDA (electrodermal activity). A potential lab-based experiment would also resolve another key limitation - the use of
an online environment. Whilst necessary to achieve sample power within budget, and whilst exclusion criteria were administered to limit distractions of phone usage and mobility, a virtual setting can’t be controlled for confounding variables such as noise, temperature and distraction which again impacts internal validity.

Alongside a potential lab experiment, replication in a natural field setting (a workplace) with a more ethnically diverse sample is also necessary to address the limitations of generalisability in this research. Ideally this would be conducted as a longitudinal study to also overcome the limitation of unknown effect adaptation over time. It would also be beneficial to focus on adjacent populations that undergo effortful tasks, such as the impact TS has on teenager affect in an effortful education environment. Finally, as trait affect - a constant individual variance that acts as an “affective lens” through which we experience the world (Barsade & Gibson, 2007) - influences the way we react to effortful situations, examining the interaction effects between trait affect, TS and PA/NA from effortful tasks would also be recommended.

### 6.2 Does framing task significance intrinsically improve affect or productivity during effortful tasks more than framing it extrinsically?

When examining this second research question, the findings are more complex. Firstly, H₃ was not supported, in that neither Extrinsic nor Intrinsic TS treatments were found to increase productivity, which is surprising as productivity is the main benefit established in the literature (e.g., Ariely et al., 2008; Grant, 2008) thus a replication was expected. This might be explained by limitations of the slider-sorting task. Whilst several pilots were run to calibrate the balance between effort and attrition, perhaps it remained too tedious to motivate quality and quantity delivery. Alternatively, we could look to the only other online experiment to replicate the effects of TS on productivity (Chandler & Kapelner, 2013) for potential explanations. This study asked participants in the meaningful treatment to sort cancerous-cell data, an inherently meaningful task connected to the TS, compared to the inherently mundane, disconnected task of
sorting sliders. Future research would therefore benefit from replicating a similarly meaningful task that is linked to the TS frame, measuring both productivity and affect, and conducting the moderation of process analysis intended for H4, to statistically examine if affect is a hidden mechanism through which TS travels.

Secondly, whilst Intrinsic TS had no significant effect on affect, results suggest that Extrinsic TS did, specifically increasing PA, (Extrinsic PA=+3.152, SE=1.284, p-value=0.015, rwolfp=0.086, d=3.273) suggesting variance between the two intervention frames, supporting H2. Again, caution must be applied, due to lost significance when testing for 9x multiple hypotheses, however, to put a real-world value on this potential effect, workers who understand the extrinsically framed task significance feel 12% increased PA following effortful work, compared to those who don’t. To put that effect into further context, Extrinsic TS interventions have the potential to shift participants from below UK population momentary PA average to above average PA (UK Population PA average=29.7 [SD=7.9]; Control PA average=27.643 [SD=9.034]; Extrinsic TS PA average=30.962 [SE=10.163]) (Boyle, 2015).

However, surprisingly, these results are not in the direction hypothesised, in that contrary to Self Determination Theory (Ryan & Deci, 1985), it is Extrinsic, rather than Intrinsic TS, that is driving the effect. Several explanations might be offered to rationalise this. One potential explanation is that the results were invalidated by limitations in the treatment design. Whilst efforts were made to make the disclosive writing questions in Intrinsic and Extrinsic treatments identical bar a few words, the Extrinsic treatment required longer contextual reading, introducing variance that may have compromised the method. If we assume internal validity however, another explanation could be found if we consider the increased cognitive load, and thus emotional load (Van Dillen et al., 2009), that the act of meaning-making demands, versus the act of consuming a meaning that someone external has already made for you – essentially, do we sweat it less when it is someone else’s ‘why’? Or perhaps we
can turn to the theory of Cognitive Dissonance (Festinger, 1962) for an explanation. Perhaps having initially connected to more personally purposeful beliefs about the task significance, the Intrinsic TS group later experienced more conflict between these beliefs and the tedious, tough experience of slider-sorting, compared to the Extrinsic TS group, ultimately creating greater cognitive discomfort and reduced capacity for PA.

A final explanation could be found in examining Vallacher & Wegner’s (1987) 3rd principle of Action Identification Theory, in which they found that the higher the level of meaning, the more resistant to change we are, and that, if we then confront difficulties, we tend to shift down levels because the lower our level of meaning, the less fixed on our beliefs we are, and the easier we find it to adapt. We could posit that Intrinsic TS sits at a higher meaning level than Extrinsic TS. If the difficult slider task naturally prompted participants to shift down levels, the Intrinsic TS group had further to travel - further meaning to effectively relinquish - which may have moderated their PA score. Simply put, is it easier to just focus on driving to the grocery shop, rather than think about nourishing your growing family, particularly when you find yourself faced with a challenge like a traffic jam? In this, a finer grained examination into the purpose/pressure thresholds (i.e., when exactly does too much purpose become too much?) would also be recommended for future research.

Overall, the suggestion of variance between the Extrinsic and Intrinsic TS frame, but also their possible interdependence discussed above, exposes potential new dimensions when we think about optimal TS formats. Perhaps truly effective interventions require us to consider a trinity of task significance sources, proposed as a model in Figure 6.2 below, drawing from 1) Extrinsic significance, sourced from objective, external understanding of task impact on others, (“I understand the big picture purpose here is to accelerate cancer research”), 2) Intrinsic significance, sourced from subjective, internal connection between extrinsic purpose and personal motivation (“my cousin is struggling with cancer – I want to do what I can to contribute to the cause”) and 3) Applied significance, sourced from recognising a connection
between the actual task composition and the extrinsic and intrinsic sources, ("whilst tedious, I can see how labelling these tumour images does its small bit to help deliver the purpose"). This model is offered to help inform future research into the TS ingredients required for maximum intervention effects.

*Figure 6.2: Trinity of Task Significance Sources*
7.0 CONCLUSION

This investigation sought to examine two critical gaps in the task significance domain of purposeful work literature by 1) exploring the role work purpose has on the emotional component of performing effortful tasks, and 2) examining extrinsic versus intrinsic work purpose via different TS interventions. Whilst the results must be handled with caution, as significance at 0.05 does not withstand correction for 9x multiple hypotheses, this study reveals a strong indication of an empirical link between task significance interventions and increased positive affect, which is a key determinate of unique workplace benefits, from enhanced executive function to increased psychological health (Loewenstein & Lerner, 2003; Yoon et al., 2022). Furthermore, this study suggests advances in our understanding of optimum TS formats by firstly indicating wellbeing effects derive from extrinsic over intrinsic significance frames, and secondly by offering the Trinity of Significance Sources model to help inform future exploration into TS intervention ingredients (Figure 6.2).

Whilst this study cautiously offers additions to the TS literature, should future research replicate and increase the generalisability of the results, potential implications could be considered within both the workplace and the classroom - two domains where effortful tasks have an everyday baring on wellbeing, and two domains where mental health issues are fast rising. Potential consequences could span how managers and teachers are trained to engage with people on tasks, how task success is appraised, how leaders lead, and how workplace and classroom cultures are designed at a macro level to draw a more clear and salient thread between the "what" and the "why".

When we recognise the fact that work is the 3rd most important driver of our total happiness (Van Praag et al., 2003), and yet also face the reality that 1 in 3 people in the UK feel unhappy at work (De Neve, 2018), the imperative to do more to address this crisis becomes compelling. Whilst it is widely acknowledged that purpose and pleasure don’t always go hand in hand - with the act of parenting as an often-cited example of
an inverse relationship (Layard, 2020) - this study challenges that paradox in the context of work, shedding new light on a potentially powerful association between knowing “why” we are performing a hard task, and the emotional pleasure we experience from it.
8.0 REFERENCES


Choi, W., Clark, J., Presslee, A., Thank, W., Bloomfield, R., Durney, M., Guggenmos, R., Libby, B., Merkley, K., Paz, M., Renes, S., Rennekamp, K., Sethuraman, M.,


Loewenstein, & Lerner. (2003). *The role of affect in decision making.*


Mccrae, R. R., John, O. P., Bond, M., Borkenau, P., Buss, D., Costa, P., Fiske, D.,
Model and Its Applications.
McKnight, P. E., & Kashdan, T. B. (2009). Purpose in Life as a System That Creates and
Sustains Health and Well-Being: An Integrative, Testable Theory. Review of
General Psychology, 13(3), 242–251. https://doi.org/10.1037/a0017152
Framing on Issue Interpretation and Risk Taking. Organisational Behaviour and
Meaning Making and Its Effects on Adjustment to Stressful Life Events.
Educational Publishing Foundation (Vol. 1, Issue 2).
Decision-Making: The Moderating Role of Interpersonal Spin. Career Development
platforms and panels for online behavioral research. Behavior Research Methods,
Understanding of Inhibition and Disease (Vol. 95, Issue 3).
Pennebaker, J. W., Kiecolt-Glaser, J. K., Glaser, R., Brumelow, S., Gordon, S., Czajka, J.,
Ferrara, K., Williams, H., Patei, H., Richards, B., Alexander, D., Cole, R., Tiebout, J.,


https://www.researchgate.net/publication/303052224


Waterman, A. (1993). *Two Conceptions of Happiness: Contrasts of Personal Expressiveness (Eudaimonia) and Hedonic Enjoyment*.


9.0 APPENDICES
9.2 Ethical Guidelines & Data Safeguarding

Having carefully consulted the LSE’s Research and Ethics Policy and Procedures, the Government’s Data Ethics Framework and the FORGOOD framework (Gill & Prowse, 2011; Lades & Delaney, 2020). I can conclude that significant ethical dimensions were considered when designing the experiment and the methods did not involve any form of deception. The TS interventions built truthful foundations whilst making sure the specific details of the treatment itself were not revealed.

The procedures were overt and transparent. With regards to additional safeguards, as the participant pool was sourced via Prolific, the experiment had the benefit of being bound by the rigorous privacy policies and procedures in place on the Prolific platform. Headlines include: “Prolific is committed to safeguarding the privacy of our website visitors, service users, and other individuals with whom we deal. As part of this commitment, we’ve ensured that we provide full privacy notices meeting the high standards of the European data protection law, known as the General Data Protection Regulation (GDPR). This summary gives you an overview, but you should read those notices (and the explanatory content linked in those notices) for fuller information.” Specifically, Prolific did not identify participants to the researcher, but simply referred only to Prolific IDs which are a string of numbers and letters. I in turn did not attempt to re-identify participants or ask for any direct identifiers like name or address.

I also provided additional privacy notices within the Qualtrics survey launched through the platform and ensured compliance with the legal obligations as data controller. In addition to this, Prolific’s security systems protect participant data and confidentiality in the following ways: “Prolific uses encrypted HTTPS connections, secured by Transport Layer Security (TLS). Participants are fully anonymized. They are assigned a unique participant ID (24 character alphanumeric). Prolific provides an anonymized internal messaging service, which allows participants to message researchers (and vice versa) with any concerns. Researchers cannot access participants’
Finally, a Data Management Plan was designed in advance of data collection and followed rigorously throughout statistical exploration and at the close of analysis.

9.3 Sample Size Justification and Realisation

A range of studies were assessed as possible reference points for effect size, however several design characteristics made them inappropriate, such as the natural field setting of the study (Salamone & Lordan, 2022) or the lack of foundational data required for sample calculations (Chandler & Kapelner, 2013). With these restrictions in mind, results from (Grant, 2008) most closely reflect the experimental procedure used in this study as they examine the effect of task purpose, in the form of stories communicating task social benefits on the productivity of callers (n=33) soliciting alumni contributions to their university. However, a large effect size is revealed at 0.95 (G-Power computation: control mean pledges earned=9.08 [SD=6.93]; treatment mean pledges earned =23:00 [SD=11.39] at alpha:0.05 and beta:0.80). This is potentially due to the experiment’s natural field setting, where the actual task’ might feel innately more meaningful as it is the individual’s real profession. As this experiment adopted similar procedures but applied them to an online setting and used a more meaningless real-effort task, it felt negligent to power for a similarly large effect size. G-Power was therefore used to compute a more conservative sample, assuming a medium effect size (d=0.5), alongside customary statistical significance (alpha=0.05) and power (beta=0.80), revealing total n=315 (treatment group 1=105, treatment group 2=105, group 3=105). Whilst it is standard to assume small effect sizes within the field of behavioural science, ignoring Grant’s large effect sizes completely and powering the study for a small effect (alpha=0.05, beta=0.08, d=0.2: n=1302) feels negligently conservative, and risked overpowering the RCT or presenting a budgetary challenge that would prevent the investigation.
In conducting the RCT, a total of 321 responses were collected in wave 1, all of whom agreed to informed consent and passed attention and seriousness checks. Participants using a mobile phone or tablet (n=6) were excluded as survey functionality required desktop or laptop devices. Of the 315 total, Condition 0: Control n=112, Condition 1: Extrinsic Task Significance n=90, Condition 2: Intrinsic Task Significance n=113. To ensure 100% of the minimum required sample was obtained for Condition 1 (n=105), Prolific was used to recruit an additional wave 2 of 15. All these additional 15 used the required devices, agreed to Informed Consent and passed attention and seriousness checks. This resulted in final group totals of Condition 1 n=112, Condition 2 n=105 and Condition 3 n=113, with an overall sample total n=330 available for analysis.

9.4 **Multiple Hypotheses Correction Testing**

Due to the multiple hypotheses represented in the framework for statistical analysis, Romano Wolf Multiple Hypothesis correction was administered to control the Family Wise Error Rate – the probability of revealing false findings or Type I errors when conducting more than one hypothesis test – and allow for dependence among p-values by bootstrap resampling (Clarke et al., 2019). In Section 5.0 Results, where statistical significance was found in standard p-values, Romano Wolf step-down adjusted p-values (rwolfp) were reported alongside them. **Table 9.4** represents the complete set of Romano Wolf step-down adjusted p-values for all 9 hypothesis tests represented in the results summarised in **Table 5.1**.
**Table 9.4.1: Romano Wolf Multiple Hypothesis Corrected P-Values For Simple Linear Regressions**

<table>
<thead>
<tr>
<th></th>
<th>Model p-value</th>
<th>Resample p-value</th>
<th>Romano-Wolf p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>papost</td>
<td>0.8116</td>
<td>0.8133</td>
<td>0.1030</td>
</tr>
<tr>
<td>napost</td>
<td>0.3353</td>
<td>0.3555</td>
<td>0.7442</td>
</tr>
<tr>
<td>scores</td>
<td>0.4574</td>
<td>0.4651</td>
<td>0.7442</td>
</tr>
<tr>
<td>papost</td>
<td>0.2429</td>
<td>0.2458</td>
<td>0.6777</td>
</tr>
<tr>
<td>napost</td>
<td>0.8149</td>
<td>0.8408</td>
<td>0.1262</td>
</tr>
<tr>
<td>scores</td>
<td>0.7138</td>
<td>0.7209</td>
<td>0.8538</td>
</tr>
<tr>
<td>papost</td>
<td>0.8360</td>
<td>0.8399</td>
<td>0.1561</td>
</tr>
<tr>
<td>napost</td>
<td>0.8513</td>
<td>0.8365</td>
<td>0.1060</td>
</tr>
<tr>
<td>scores</td>
<td>0.7191</td>
<td>0.6977</td>
<td>0.8538</td>
</tr>
</tbody>
</table>

**Table 9.4.2: Romano Wolf Multiple Hypothesis Corrected P-Values For Multiple Linear Regressions**

<table>
<thead>
<tr>
<th></th>
<th>Model p-value</th>
<th>Resample p-value</th>
<th>Romano-Wolf p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>papost</td>
<td>0.0394</td>
<td>0.0332</td>
<td>0.1761</td>
</tr>
<tr>
<td>napost</td>
<td>0.0346</td>
<td>0.0366</td>
<td>0.1061</td>
</tr>
<tr>
<td>scores</td>
<td>0.6208</td>
<td>0.6379</td>
<td>0.7940</td>
</tr>
<tr>
<td>papost</td>
<td>0.0179</td>
<td>0.0199</td>
<td>0.8864</td>
</tr>
<tr>
<td>napost</td>
<td>0.2969</td>
<td>0.2857</td>
<td>0.6445</td>
</tr>
<tr>
<td>scores</td>
<td>0.5447</td>
<td>0.5748</td>
<td>0.7940</td>
</tr>
<tr>
<td>papost</td>
<td>0.2307</td>
<td>0.2591</td>
<td>0.6445</td>
</tr>
<tr>
<td>napost</td>
<td>0.0100</td>
<td>0.0332</td>
<td>0.8664</td>
</tr>
<tr>
<td>scores</td>
<td>0.7563</td>
<td>0.7887</td>
<td>0.7940</td>
</tr>
</tbody>
</table>
9.5 Exploratory Analyses

9.5.1 Exploratory Analyses Overview

Exploratory analysis was conducted on DV “wta” – willingness to accept payment to repeat the task, detailed below. It is also worth noting that in preregistration a plan to explore the interaction effects of Personality was outlined, however the Big 5 Personality questionnaire was removed due to budgetary constraints.

9.5.2 Exploratory Analyses – DV3 “wta”

Total RCT (n=330) mean “wta” measured in £0.000 is 5.649 (SD=4.272), with a right skew (2.017) and mild kurtosis (7.091). Median “wta” is 4.090, with a range of 0.000 – 20.000. Extrinsic TS group (n=105) has lower mean wta 5.304 (SD=4.124) than Intrinsic TS group (n=113) 5.658 (SD=3.973), and both have lower mean wta compared to control group (n=112) 5.965 (SD=4.693).

Two tailed t-tests reveal that, compared to control group, ETA wta mean is 0.661 (SE=0.601) higher and Intrinsic TS wta mean is 0.307 (SE=0.581) higher, however neither were significant (p-values=>0.05) therefore we cannot reject the null that there is no treatment effect on wta. In comparing scores between Extrinsic TS and Intrinsic TS groups, again no significant findings were found. ANOVA tables uphold these findings, as do simple and multiple linear regression analysis (with and without outliers) and Mann Whitney tests, conducted for robustness checks (all p-values=>0.05). Therefore, we cannot reject the null hypothesis that there is no treatment effect on willingness to accept payment to repeat the task.
9.6  Survey Detail

9.6.1  Original Slider-Sorting Design Reference (Gill & Prowse, 2011)

*Figure 9.6* depicts the original real-effort task used in Gill & Prowse’s (2011) study where individuals sorted sliders into order on computers within a laboratory setting.

*Figure 9.6: Original Slider Task Schematic*

Slider schematic showing task pre and post completion i.e. the slider has successfully been moved from 0 to 50, and example screen displayed to participants on laboratory computer. Both taken from (Gill & Prowse, 2011)