Debiasing Optimistic Planning in Knowledge Work: A Human-Centered Approach

YOANA AHMETOGLU

Primary supervisor: **Prof Duncan Brumby**

Secondary supervisor: **Prof Anna Cox**

Examining committee: **Prof Adam Harris and Dr John Rooksby**

Word count: **56 133**

Thesis submission date: **30 September 2025**

Submitted to University College London (UCL) in partial fulfilment of the requirements for the award of the degree of Doctor of Philosophy.

Declaration

I, **Yoana Ahmetoglu**, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

[redacted]

Yoana Ahmetoglu

London, United Kingdom

30 September 2025

Abstract

Task planning – deciding which tasks to complete and when – is essential for productivity in academia and beyond, yet individuals often struggle to make accurate plans: those they can finish on schedule and in roughly the expected time. One contributor to this challenge is the optimistic planning bias: the tendency to misestimate how long tasks will take, which often results in missed deadlines and work stress. Despite its wide scale and impact, planning accuracy remains an underexplored design concept in Human-Computer Interaction (HCI). This thesis investigates how technologies can support task planning by translating psychological theories of bias mitigation into HCI interventions.

The studies use a range of empirical and review methods. Study 1 examines why early career researchers struggle with planning accurately, identifying the optimistic planning bias as a key reason for inaccuracies. Study 2 builds on this by examining how a similar group of early career researchers attempted to plan more accurately during the Covid-19 lockdown, identifying strategies such as task breakdown and manual time tracking. Study 3 extends these findings by synthesising psychological literature to identify four evidence-based debiasing strategies: duration feedback, distributional data, task breakdown, and induced neutrality. These strategies overlap with, but formalise and expand on, those observed in Study 2. Study 4 evaluates 47 commercial task management tools, showing that support for these strategies is limited or inconsistent. Building on this foundation, Studies 5 and 6 focus on duration feedback, a strategy observed in practice and well supported in theory, yet still underrepresented in PTM tools. Two field interventions with postgraduate and doctoral students explore how duration feedback can be utilised to improve accuracy of plans in real-world academic settings and how accuracy-oriented planning technologies can be designed in human-centered ways.

This thesis contributes to HCI by extending our understanding of planning as a lived practice shaped by optimistic bias, by bridging psychology and HCI through design-relevant constructs and by demonstrating that the impact of debiasing interventions lies in shaping the experience of planning as well as in improving accuracy.

Impact Statement

Within research, this thesis has advanced understanding of optimistic planning bias in real-world contexts, made psychological theories of debiasing accessible for design, and shown how such strategies can be operationalised and tested in HCI studies, thereby shaping future research on planning technologies. The research has been disseminated through peer-reviewed publications at CSCW, IEEE, and CHIWORK, through participation in doctoral consortia (ECSCW and CHIWORK), and through presentations at industry events such as a Microsoft symposium. I also contributed to community engagement around this topic by co-organising a CHIWORK'25 workshop on AI and wellbeing at work, which built on themes from this thesis and extended their discussion within the HCI community. Beyond HCI, the findings contribute to psychology by extending research on the time estimation bias into naturalistic settings, and to organisational studies by highlighting planning accuracy as a factor in workload management.

Beyond academic research, this thesis demonstrates how optimistic planning bias affects the everyday work of students and staff, providing evidence that can inform workload policies and support systems. Because inaccurate planning contributes to stress and burnout, particularly in knowledge-intensive professions with heavy workloads, this research supports the development of planning tools that help people align intended work with actual capacity. More broadly, the thesis advances public understanding of cognitive bias by showing how optimistic planning shapes everyday work and how the design of digital tools can affect the way such biases are experienced and managed.

Publications and Research Declaration Forms

Full Papers

1. **Ahmetoglu, Y.**, Iskandar, A., Taoh, S., Ying, A., Brumby, D. P., & Cox, A. L. (2025). "I work much better by doing less": How task duration feedback affects optimistic planning bias in academic work. *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work (CHIWORK '25)*. ACM.

2. **Ahmetoglu, Y.**, Brumby, D., & Cox, A. (2024, June). Bridging the gap between time management research and task management app design: A study on the integration of planning fallacy mitigation strategies. In *Proceedings of the 3rd Annual Symposium on Human-Computer Interaction for Work* (pp. 114). ACM.

3. **Ahmetoglu, Y.**, Brumby, D. P., & Cox, A. L. (2021). Disengaged from planning during the lockdown? An interview study in an academic setting. *IEEE Pervasive Computing*, *20*(4), 1825.

4. **Ahmetoglu, Y.**, Brumby, D. P., & Cox, A. L. (2021). To plan or not to plan? A mixed-methods diary study examining when, how and why knowledge work planning is inaccurate. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW3), 120. **Honourable Mention Award**

Workshop Papers, Late-Breaking Work, and Position Papers

1. **Ahmetoglu, Y.**, Somanath, S., Lallemand, C., Solovey, E. T., Brumby, D. P., and Cox, A. L. (2025, June). Paving the Way for AI that Supports Flourishing at Work. In Adjunct Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work (pp. 1-3).

2. **Ahmetoglu, Y.**, Brumby, D. P., & Cox, A. L. (2020). Time estimation bias in knowledge work: Tasks with fewer time constraints are more error-prone. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*.

3. **Ahmetoglu, Y.** (2020). Understanding and supporting daily planning in knowledge work. In *European Society for Socially Embedded Technologies (EUSSET)*.

4. **Ahmetoglu, Y.**, Brumby, D., & Cox, A. (2020). A longitudinal interview study on work planning during COVID-19 lockdown. *Microsoft*.

5. **Petrova, Y.**, Brumby, D. P., & Cox, A. L. (2019). Future challenges in design for multitasking at work. In *CHI 2019 Workshop: Technology to Mediate Role Conflict in Motherhood*.

6. **Petrova, Y.**, Brumby, D. P., & Cox, A. L. (2019). Motherhood and employment: Challenges for HCI research on work-life conflict. Position paper submitted to *CHI 2019 Workshop: Technology to Mediate Role Conflict in Motherhood*.

UCL Research Paper Declaration Form

Referencing the Doctoral Candidate’s Own Published Work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server
- are in submission to a peer-reviewed publication
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you will complete this form twice.

1. **For a research manuscript that has already been published**
   (if not yet published, skip to section 2):

   a) **Title of the manuscript:**

   To plan or not to plan? A mixed-methods diary study examining when, how and why knowledge work planning is inaccurate

   b) **Provide the DOI or direct link to the published the work:**

   https://doi.org/10.1145/3432921

   c) **Publication name (e.g. journal or textbook):**

   Proceedings of the ACM on Human-Computer Interaction, Volume 4, Issue CSCW3

   d) **Publisher name (e.g. Elsevier, Oxford University Press):**

   ACM Journals

   e) **Date of publication:**

   05 January 2021

   f) **List all authors as they appear in the publication:**

   Yoana Ahmetoglu, Duncan P. Brumby, Anna L. Cox

   g) **Was the work peer reviewed?**

   Yes

   h) **Do you retain copyright for the work?**

   Yes

   i) **Was an earlier version uploaded to a preprint server (e.g., medRxiv, arXiv)?**
   If ‘Yes’, provide the DOI or direct link. If not applicable, leave blank.
   https://discovery.ucl.ac.uk/id/eprint/10126354/7/Brumby_CSCW_UCL_OA_1.pdf

   If ‘No’, please seek publisher permission and check the box below:

   ☐

   I acknowledge permission of the publisher named in item **1d** to include in this thesis portions of the publication cited in item **1c**.

2. **For a manuscript prepared for publication but not yet published**
(if already published, skip to section 3):

a) **Current title of the manuscript:**

Click or tap here to enter text.

b) **Has it been uploaded to a preprint server** (e.g., medRxiv, arXiv)**?**
If 'Yes', provide the DOI or direct link. If not applicable, leave blank.

Click or tap here to enter text.

c) **Intended publication outlet (e.g., journal name):**

Click or tap here to enter text.

d) **List all authors in the intended authorship order:**

Click or tap here to enter text.

e) **Current stage of publication (e.g., in submission, under review):**

Click or tap here to enter text.

For multi-authored work, please provide a contribution statement detailing each author's role (if single-authored, skip to section 4):
All of the papers included in this thesis were co-authored with my supervisors. I was the lead researcher on each study, responsible for conceiving the research questions, designing the studies, collecting and analysing the data, and writing the first drafts of the manuscripts. My supervisors contributed through regular guidance and feedback, advising on study design, supporting interpretation of findings, and providing editorial input during the revision process. Final versions of the papers were developed collaboratively, with me as first author.

3. **In which chapter(s) of your thesis can this material be found?**

Chapter 3

4. **e-Signatures confirming accuracy of the above information**
(This form should be co-signed by the supervisor/senior author unless the work is single-authored)

Candidate signature:

Yoana Ahmetoglu

Date:

23.09.2025

Supervisor/ Senior Author (where appropriate):

Duncan Brumby

Date:

23 September 2025

UCL Research Paper Declaration Form

Referencing the Doctoral Candidate's Own Published Work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server
- are in submission to a peer-reviewed publication
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you will complete this form twice.

1. **For a research manuscript that has already been published**
(if not yet published, skip to section 2):

a) **Title of the manuscript:**

Disengaged From Planning During the Lockdown? An Interview Study in an Academic Setting

b) **Provide the DOI or direct link to the published the work:**

https://doi.org/10.1109/MPRV.2021.3118900

c) **Publication name (e.g. journal or textbook):**

IEEE Pervasive Computing

d) **Publisher name (e.g. Elsevier, Oxford University Press):**

IEEE

e) **Date of publication:**

28 October 2021

f) **List all authors as they appear in the publication:**

Yoana Ahmetoglu, Duncan P. Brumby, Anna L. Cox

g) **Was the work peer reviewed?**

Yes

h) **Do you retain copyright for the work?**

Yes

i) **Was an earlier version uploaded to a preprint server (e.g., medRxiv, arXiv)?**
If 'Yes', provide the DOI or direct link. If not applicable, leave blank.
https://discovery.ucl.ac.uk/id/eprint/10138896/

If 'No', please seek publisher permission and check the box below:

☐

I acknowledge permission of the publisher named in item **1d** to include in this thesis portions of the publication cited in item **1c**.

2. **For a manuscript prepared for publication but not yet published**
(if already published, skip to section 3):

a) **Current title of the manuscript:**

Click or tap here to enter text.

b) **Has it been uploaded to a preprint server** (e.g., medRxiv, arXiv)**?**
If 'Yes', provide the DOI or direct link. If not applicable, leave blank.

Click or tap here to enter text.

c) **Intended publication outlet (e.g., journal name):**

Click or tap here to enter text.

d) **List all authors in the intended authorship order:**

Click or tap here to enter text.

e) **Current stage of publication (e.g., in submission, under review):**

Click or tap here to enter text.

For multi-authored work, please provide a contribution statement detailing each author's role (if single-authored, skip to section 4):
All of the papers included in this thesis were co-authored with my supervisors. I was the lead researcher on each study, responsible for conceiving the research questions, designing the studies, collecting and analysing the data, and writing the first drafts of the manuscripts. My supervisors contributed through regular guidance and feedback, advising on study design, supporting interpretation of findings, and providing editorial input during the revision process. Final versions of the papers were developed collaboratively, with me as first author.

3. In which chapter(s) of your thesis can this material be found?

Chapter 3

4. e-Signatures confirming accuracy of the above information
(This form should be co-signed by the supervisor/senior author unless the work is single-authored)

Candidate signature:

Yoana Ahmetoglu

Date:

23.09.2025

Supervisor/ Senior Author (where appropriate):

Duncan Brumby

Date:

23 September 2025

UCL Research Paper Declaration Form

Referencing the Doctoral Candidate's Own Published Work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server
- are in submission to a peer-reviewed publication
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you will complete this form twice.

1. **For a research manuscript that has already been published**
   (if not yet published, skip to section 2):

   a) **Title of the manuscript:**

   Bridging the gap between time management research and task management app design: A study on the integration of planning fallacy mitigation strategies

   b) **Provide the DOI or direct link to the published the work:**

   https://doi.org/10.1145/3663384.3663404

   c) **Publication name (e.g. journal or textbook):**

   CHIWORK '24: Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work

   d) **Publisher name (e.g. Elsevier, Oxford University Press):**

   Association for Computing Machinery (ACM)

   e) **Date of publication:**

   25 June 2024

   f) **List all authors as they appear in the publication:**

   Yoana Ahmetoglu, Duncan P. Brumby, Anna L. Cox

   g) **Was the work peer reviewed?**

   Yes

   h) **Do you retain copyright for the work?**

   Yes

   i) **Was an earlier version uploaded to a preprint server (e.g., medRxiv, arXiv)?**
   If 'Yes', provide the DOI or direct link. If not applicable, leave blank.
   https://discovery.ucl.ac.uk/id/eprint/10138896/

   If 'No', please seek publisher permission and check the box below:

   ☐
   I acknowledge permission of the publisher named in item **1d** to include in this thesis portions of the publication cited in item **1c**.

2. For a manuscript prepared for publication but not yet published
(if already published, skip to section 3):

a) **Current title of the manuscript:**

Click or tap here to enter text.

b) **Has it been uploaded to a preprint server** (e.g., medRxiv, arXiv)**?**
If 'Yes', provide the DOI or direct link. If not applicable, leave blank.

Click or tap here to enter text.

c) **Intended publication outlet (e.g., journal name):**

Click or tap here to enter text.

d) **List all authors in the intended authorship order:**

Click or tap here to enter text.

e) **Current stage of publication (e.g., in submission, under review):**

Click or tap here to enter text.

For multi-authored work, please provide a contribution statement detailing each author's role (if single-authored, skip to section 4):
All of the papers included in this thesis were co-authored with my supervisors. I was the lead researcher on each study, responsible for conceiving the research questions, designing the studies, collecting and analysing the data, and writing the first drafts of the manuscripts. My supervisors contributed through regular guidance and feedback, advising on study design, supporting interpretation of findings, and providing editorial input during the revision process. Final versions of the papers were developed collaboratively, with me as first author.

3. In which chapter(s) of your thesis can this material be found?

Chapter 4

4. e-Signatures confirming accuracy of the above information
(This form should be co-signed by the supervisor/senior author unless the work is single-authored)

Candidate signature:

Yoana Ahmetoglu

Date:

23.09.2025

Supervisor/ Senior Author (where appropriate):

Duncan Brumby

Date:

23 September 2025

UCL Research Paper Declaration Form

Referencing the Doctoral Candidate's Own Published Work(s)

Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:

- have been uploaded to a preprint server
- are in submission to a peer-reviewed publication
- have been published in a peer-reviewed publication, e.g. journal, textbook.

This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you will complete this form twice.

1. For a research manuscript that has already been published
(if not yet published, skip to section 2):

a) **Title of the manuscript:**

"I Work Much Better by Doing Less": How Task Duration Feedback Affects Optimistic Planning Bias in Academic Work

b) **Provide the DOI or direct link to the published the work:**

https://doi.org/10.1145/3729176.3729182

c) **Publication name (e.g. journal or textbook):**

CHIWORK '25: Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work

d) **Publisher name (e.g. Elsevier, Oxford University Press):**

Association for Computing Machinery (ACM)

e) **Date of publication:**

22 June 2025

f) **List all authors as they appear in the publication:**

Yoana Ahmetoglu, Duncan P. Brumby, Anna L. Cox

g) **Was the work peer reviewed?**

Yes

h) **Do you retain copyright for the work?**

Yes

i) **Was an earlier version uploaded to a preprint server (e.g., medRxiv, arXiv)?**
If 'Yes', provide the DOI or direct link. If not applicable, leave blank.
https://discovery.ucl.ac.uk/id/eprint/10138896/

If 'No', please seek publisher permission and check the box below:

☐

I acknowledge permission of the publisher named in item **1d** to include in this thesis portions of the publication cited in item **1c**.

2. **For a manuscript prepared for publication but not yet published**
   (if already published, skip to section 3):

   a) **Current title of the manuscript:**

   Click or tap here to enter text.

   b) **Has it been uploaded to a preprint server** (e.g., medRxiv, arXiv)**?**
   If 'Yes', provide the DOI or direct link. If not applicable, leave blank.

   Click or tap here to enter text.

   c) **Intended publication outlet (e.g., journal name):**

   Click or tap here to enter text.

   d) **List all authors in the intended authorship order:**

   Click or tap here to enter text.

   e) **Current stage of publication (e.g., in submission, under review):**

   Click or tap here to enter text.

   For multi-authored work, please provide a contribution statement detailing each author's role (if single-authored, skip to section 4):
   All of the papers included in this thesis were co-authored with my supervisors. I was the lead researcher on each study, responsible for conceiving the research questions, designing the studies, collecting and analysing the data, and writing the first drafts of the manuscripts. My supervisors contributed through regular guidance and feedback, advising on study design, supporting interpretation of findings, and providing editorial input during the revision process. Final versions of the papers were developed collaboratively, with me as first author.

3. **In which chapter(s) of your thesis can this material be found?**

   Chapter 5

4. **e-Signatures confirming accuracy of the above information**
   (This form should be co-signed by the supervisor/senior author unless the work is single-authored)

   Candidate signature:

   Yoana Ahmetoglu

   Date:

   23.09.2025

   Supervisor/ Senior Author (where appropriate):

   Duncan Brumby

   Date:

   23 September 2025

Acknowledgements

I'd like to express my gratitude... To my supervisors, Duncan and Anna, for believing in me from day one. I've been blessed not with one, but with two incredible supervisors, mentors and guides throughout this journey. Thank you from the bottom of my heart... To my fellow students, who are too many to name one by one, thank you for the generosity, encouragement, and everyday camaraderie you have shown me. You have taught me what genuine support looks like... To my parents and grandparents, mame and tate, baba and dyado, who sacrificed literally everything they had, so I could become the first person with not only a university degree, but a doctoral one. To those still with me, and those who could not celebrate with me, you are everything to me. To my husband, Gorkan, who was always there to help, always there to listen to my worries, and always there to reassure me that what lays ahead would be just as I imagined it. Thank you for being incredibly biased about me in a very positive way. And to my children, who may one day find this thesis, and wonder what life was back then – doing this while having you was the best thing I could have wished for. Last but not least, thank you to my friends, especially Carolina, who supported me to navigate my doctoral life while becoming a mom (twice). Doing a PhD requires a village, and I was lucky to have it.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Chapter Outline

In this chapter, I introduce the problem that serves as the central motivation for this thesis. I outline the research questions, contributions, scope, and approach, while also providing an overview of the thesis structure.

This chapter introduces the problem of accurate task planning in academic knowledge work and the need to understand why plans diverge from outcomes and how digital tools might provide better support. It sets out the research questions that guided the thesis: what challenges academics face in planning tasks accurately, how existing strategies and tools address these challenges, and how technology design can better support accuracy in real-world planning. The chapter outlines the scope and approach, describing a six-study programme that combined field observations, literature and functionality reviews, and intervention trials in academic contexts. In doing so, it previews the main contributions: showing how optimistic planning bias was embedded in everyday planning practices, developing the concept of reflective planning to explain how people engaged with mismatches between plans and outcomes, and drawing implications for the design of planning technologies.

1.1 Motivation

Imagine a postdoctoral researcher, Alex, a few weeks into her new role. Eager to make progress, she sits down over breakfast and quickly notes a plan for the day: *1) manuscript revisions, 2) talk outline, 3) data check.* Alex feels motivated and confident, believing that keeping her plan flexible will be useful for staying productive throughout the day.

As the day unfolds, things do not go as expected. Manuscript revisions take longer than anticipated as Alex gets caught up refining arguments and searching for missing references. A quick check of emails turns into a lengthy back-and-forth with a collaborator requesting additional analysis. By the afternoon, outlining the talk feels daunting after a morning of interruptions. When Alex finally turns to the dataset, she is already mentally drained. As the evening approaches, Alex realizes she has only completed part of what she set out to do. Frustrated, she wonders if she is managing her time poorly or simply expecting too much. Over time, she becomes hesitant to plan in detail, wary of the recurring disappointment of unfinished tasks.

The described experience above illustrates a widespread challenge faced by many knowledge workers today, such as academics and researchers. While task planning – deciding which tasks to complete and when – is intended to provide structure, reduce work stress and serve as a source of motivation, establishing a consistent planning routine often proves difficult due to the frequent failure of plans [1, 2]. Research indicates that when individuals create plans, they leave up to 27% of their planned tasks incomplete by the end of the day, often prioritizing factors such as urgency over importance [3]. Despite the growing recognition of this issue, the underlying reasons for inaccurate planning and the most effective and beneficial strategies to support planning accuracy remain poorly understood. Existing evidence suggests that the persistent disconnect between planned tasks and actual outcomes (leaving unfinished tasks by end of the day) contributes significantly to stress, reduced job satisfaction, and poor sleep and work-life balance [4, 5]. While digital planning tools are widely adopted, they are generally presented as systems for organising and recording tasks [6]. It remains less clear whether, or how, such tools help improve the accuracy of plans (what actually gets done), and the effects of such interventions in practice.

Existing research has examined aspects of planning in different fields, but often in ways that do not connect well to everyday work. In psychology, studies of time estimation biases typically rely on controlled, short tasks, which isolate specific cognitive processes but overlook how planning unfolds across multiple, interdependent tasks in daily work [7]. Organisational research has highlighted the benefits of regular planning routines but provides little insight into the mechanisms that help people develop or sustain such habits [4]. HCI research, meanwhile, has explored planning tools and practices, but has focused largely on usability and engagement rather than how technologies might address systematic inaccuracies in planning [8, 9, 1]. There remains limited understanding of how planning accuracy is enacted in everyday academic work, and of how insights from psychology and HCI can be integrated to inform the design of planning technologies.

1.2 Research question and answer

The aim of this thesis is to investigate how digital technologies can support more accurate task planning, by combining evidence-based strategies from psychology and HCI with empirical observations of academic work, and translating these into design guidelines for human-centered planning tools. The overarching research question, therefore, of this thesis is: **In what ways can human-centered planning technologies be designed to support more accurate task planning?**

This question is addressed through three sub-questions:

1. **What are the main challenges that academic knowledge workers face in planning work tasks accurately?** As a first step, this thesis aims to gain an in-depth understanding of planning *in situ* focusing on exploring task planning inaccuracies and on obtaining individuals' perceptions and accounts about their task planning practices and needs. It identifies the optimistic planning bias as a key reason for task planning inaccuracies.

2. **To what extent do existing planning technologies support users in making accurate plans?** Optimistic bias contributes to planning failures, yet the role of personal task management (PTM) tools (i.e. to-do list apps) in mitigating this bias remains unclear. The second step in the thesis synthesizes evidence-based bias mitigation strategies to evaluate the extent to which PTM tools support the implementation of these strategies.

3. **What are the effects of task planning debiasing interventions on planning and productivity outcomes?** To explore the effectiveness of underrepresented interventions in PTM tools for increasing task planning accuracy in academic settings, the third step in this thesis examines

how the debiasing strategy of task duration feedback can be applied as a debiasing intervention, while also considering its effects on the lived experience of planning at work.

By addressing these questions, this thesis provides the following **answer** to the overarching research question: Technologies can better support accurate planning not by narrowly enforcing precision in time estimates, but by scaffolding the reflective processes through which people manage the gap between plans and outcomes. The studies showed that strategies such as task breakdown and duration feedback can help, but their value lies less in producing perfect accuracy than in supporting people to adjust expectations, increase self-efficacy, and reduce the impact of repeated deviations. Effective planning support should therefore be selective, targeting the most consequential inaccuracies rather than every small deviation, and context-sensitive, recognising that accuracy is most beneficial when individuals have sufficient control over their tasks to enact change.

1.3 Contribution

This thesis contributes to HCI by advancing timely discussions on performance-focused HCI research and the need for more human-centric understandings of productivity [10]. It introduces planning accuracy as a complementary dimension of productivity, and develops the construct of reflective planning to explain how people engage with mismatches between plans and outcomes. Together, these contributions show how tools can better support the creation of feasible plans and help mitigate the overcommitment that undermines personal task management [11].

Studies in HCI have largely approached productivity by supporting users in completing tasks efficiently or more quickly. For example, systems have been developed to help users maintain focus, avoid interruptions, or coordinate multiple tasks [12, 13]. By contrast, little attention has been given to how accurately people plan their work in the first place. Rather than focusing on maximising output, this research begins to open up a space in HCI for considering planning technologies that help people work within their realistic capacity while balancing multiple demands. In doing so, the thesis contributes to HCI by providing design knowledge for planning tools that reflect the real practices and challenges of academic knowledge workers. It provides empirical insights into the factors shaping planning accuracy, design guidelines for developing digital planning tools that can help mitigate optimistic bias, and practical evaluations of existing applications and feedback interventions to inform future tool development.

In particular, this thesis makes the following specific contributions to the field of HCI:

1. **Real-world planning challenges in academic work**: Studies 1 and 2 provide evidence on how academic knowledge workers made vague and overly optimistic daily plans and experienced disengagement from planning during periods of disruption and stress. They quantified the extent of planning bias across different task types and developed a typology of factors contributing to delays, including vague task definitions, hidden subtasks, and interruptions. The studies also show that through reflecting on their planning practices, participants adopted strategies such as task breakdowns and time tracking, which supported re-engagement with planning and were perceived as improving planning accuracy.

2. **Translation of theory-based strategies for planning tools design:** Study 3 synthesizes four debiasing strategies, bridging insights from psychology and HCI, to provide a conceptual language for examining how planning tools might address bias. These strategies include duration feedback (using past task times to guide new estimates), distributional data (drawing on multiple examples rather than a single case), task breakdown (subdividing complex activities into smaller steps), and induced neutrality (prompting estimates in a less emotionally charged frame). These strategies overlap with, but formalise and expand on, those observed in Study 2.

3. **Identification of design gaps in digital planning tools:** Study 4 conducted a functionality review of 47 commercial task management applications, evaluating the extent to which they supported evidence-based strategies for improving planning. The analysis showed that features directly addressing planning accuracy were largely absent or burdensome to use, leaving users with limited support for managing bias in practice. By mapping these shortcomings, the study identifies specific design gaps that future research and tool development can address to better support realistic planning.

4. **Empirical insights into the limitations and potential of feedback for planning accuracy:** Studies 5 and 6 introduced duration feedback into academic work contexts, providing one of the first empirical tests of debiasing strategies in real-world planning. Across both studies, feedback did not lead to reliable improvements in minute-level accuracy at the group level. In Study 6, however, diary data showed a shift in how participants evaluated their plans: optimistic classifications decreased and realistic classifications increased. Participants also reported reductions

in procrastination and greater self-efficacy, complemented by rich qualitative insights. These studies contribute by identifying how task duration feedback might be embedded in planning tools to support everyday use.

5. **Design implications for planning tools to support accurate planning:** The thesis synthesises empirical and theoretical insights into how optimistic planning bias is experienced in practice and develops design implications for re-thinking digital planning support. The findings highlight that tools should scaffold reflection, recalibration, and engagement with planning, rather than focusing narrowly on error correction or task organization. The General Discussion integrates evidence across all six studies to outline directions for future research and tool development.

1.4 Research approach

This thesis adopts a mixed-methods approach combining observational, review, and intervention methods. The research proceeded in a sequenced way: (1) observing and characterising everyday planning, (2) reviewing psychological evidence to identify mechanisms to improve accuracy of dialy plans, and (3) testing those mechanisms through field interventions. Following Bryman [14], mixed methods served four functions: triangulation, complementarity, development, and expansion.

1.4.1 Observational Studies (Studies 1 and 2)

Study 1 used a diary and interview design adapted from Newman et al.'s augmented diary method [15]. Participants completed morning planning entries (tasks and estimated durations) and evening logs (actual start/end times), enabling direct comparison between intended and actual durations at the task level. This approach was selected over automated time-tracking because only diary methods capture intended plans alongside outcomes, allowing optimistic bias to be measured at the user-defined task level. Semi-structured interviews explored reasons for discrepancies and the role of interruptions and shifting priorities. A lab experiment would have offered less ecological validity and would not have captured these lived contingencies. Study 2 extended this work using longitudinal interviews during COVID-19 lockdowns. Repeated interviews foregrounded how planning practices changed under disruption and captured the emergence of strategies over time (e.g., manual time tracking, task breakdown).

1.4.2 Translating Theory into Designable Constructs (Studies 3 and 4)

Study 3 conducted a rapid systematic review of psychology experiments on debiasing time estimation. Controlled experiments in psychology have identified mechanisms such as feedback and distributional information, but these findings are scattered and typically use tasks far removed from knowledge-work contexts. A structured search and eligibility process based on PRISMA [16] ensured transparency while keeping the scope feasible for the thesis. The output was four strategies (duration feedback, distributional data, task breakdown, induced neutrality), and it served as a design-ready framework. Less structured narrative reviews risked selective bias, while a full meta-analysis would have obscured the translational purpose of the review. Study 4 used these strategies to conduct a functionality review of 47 commercial task-management apps. Functionality reviews [17] offer a systematic way to assess whether digital tools implement evidence-based needs. Using inter-rater checks for reliability, the analysis mapped how apps supported (or failed to support) the four strategies. User surveys or app-store analyses were avoided because they would not reveal design-level implementation or allow theory-driven mapping at feature level.

1.4.3 Intervention Studies (Studies 5 and 6)

Study 5 examined the feasibility of duration feedback in everyday academic work using *Sunsama*, a commercial planning tool integrating duration estimation and time-tracking. It provided an ecologically valid environment to observe how duration feedback is adopted within existing routines. However, participants used these features inconsistently, motivating a simpler approach. Study 6 therefore employed a researcher-designed spreadsheet tracker requiring manual time logging against self-defined tasks. Although effortful, it offered a simple, transparent, and controllable method for capturing planned versus actual durations. Automated trackers were unsuitable because they log software usage rather than user-defined tasks. Both studies combined behavioural logs, pre-/post-surveys (optimism, procrastination, self-efficacy), and qualitative interviews. This triangulation enabled examination not only of whether feedback worked but how participants engaged with it and under what conditions it improved planning accuracy.

1.4.4 Ethical Considerations

All studies received approval from the UCL Research Ethics Committee. Participants were informed about the purpose of each study, the nature of the tasks involved, and their right to withdraw at any time without giving a reason. Because task lists and planning diaries sometimes contained personal or sensitive information, all data were anonymised and stored securely, and only the research team had access to the raw materials. For the intervention studies, participants gave explicit consent to use a commercial tool such as Sunsama, and no private account information or workspace content was collected. The diary and time-tracking requirements created some burden, so participants were informed of the expected commitment before taking part and were compensated appropriately. During the COVID-19 lockdown study, interviews were conducted remotely and care was taken to acknowledge the stress and disruption participants were experiencing. Across all studies, the emphasis was on transparency, minimising burden, protecting privacy, and ensuring that participation was voluntary and conducted with respect for participants' wellbeing.

1.5 Research scope

This thesis is situated at the intersection of knowledge work contexts and digital planning technologies. The scope is defined along two dimensions: the work setting of academic knowledge workers (specifically, early career researchers) and the technological domain of personal task management (PTM) tools.

Early Career Researchers in Academia

Academic work provides a particularly relevant context for examining planning accuracy and the optimistic planning bias, especially for early career researchers (ECRs). Academic roles combine high autonomy with ambiguous goals, shifting expectations, and long-term self-directed projects, conditions that are well known to increase the likelihood of unrealistic planning. Teaching, supervision, and administrative demands often arise unpredictably and reshape workdays at short notice, undermining carefully constructed plans and intensifying workloads [9, 2]. These characteristics make academia a strong setting for observing how planning failures develop and how technological interventions might help mitigate them.

Systemic pressures heighten the stakes of inaccurate planning. ECRs frequently report long working hours, job insecurity, rising performance expectations, and substantial administrative load, all of which contribute to stress, burnout, and worklife conflict [11]. Within this environment, planning accuracy is not merely a cognitive challenge but also a wellbeing concern: unrealistic plans accumulate into overload, disrupted rest, and persistent dissatisfaction. Cultural norms that valorise overcommitment further reinforce biases, particularly for cognitively demanding tasks such as writing, reviewing, and grant preparation [18]. For ECRs, whose work combines long-term research with immediate teaching and service obligations, the need for accurate and adaptive planning is both essential and consistently difficult to meet.

Although this thesis focuses on ECRs in academia, the issues it addresses are relevant across many knowledge-intensive professions. Software developers must coordinate deep problem-solving with reactive maintenance [19]; consultants constantly rebalance priorities across clients; R&D professionals respond to timelines shaped by discovery and funding cycles [3]; and healthcare coordinators must manage heavy administrative responsibilities alongside research pressures. These fields have been studied extensively in relation to workload and burnout [20, 21, 22], yet planning accuracy itself has received comparatively little attention. By analysing planning failures within academia, this thesis generates insights that extend to other domains where high autonomy, fragmented demands, and uncertainty shape everyday work.

To support this investigation, the studies in this thesis draw on different populations and methodological approaches. Studies 1 and 2 use mixed samples of ECRs in observational designs to capture everyday planning practices. Studies 3 and 4 do not involve participants, as they examine psychological literature and commercial task management tools, including those commonly reported by participants in the observational studies (e.g., Notion, Trello). Studies 5 and 6 are field interventions where greater sample homogeneity was required: Study 5 involved postgraduate students in the coursework phase of their programmes, whose work patterns shared features with academic planning; Study 6 focused exclusively on PhD students to ensure comparability during the in-the-field trial.

Personal Task Management Tools

This thesis examines personal task management (PTM) tools, which are technologies that help individuals capture, organise, and keep track of their tasks across time. PTM tools include a variety of applications such as to-do lists, calendars, and hybrid systems that combine both. Within this broader category, the primary focus of the thesis is on digital to-do list applications. To-do lists explicitly support the everyday act of deciding what to do and when to do it, making them a natural site for investigating planning accuracy. Unlike calendars, which often reflect externally imposed constraints such as meetings or fixed deadlines, to-do lists are primarily user-driven: individuals must decide what to include, how to structure it, and when to act. This makes them especially revealing of the cognitive strategies, assumptions, and biases that underlie planning decisions. To-do list apps also vary widely in the degree of structure they impose (from priority lists to open-ended workspaces) and are used across diverse knowledge work contexts. This flexibility highlights both their potential and their limitations in supporting realistic planning.

Calendars are also PTM tools, but they are not the central focus of this thesis. They appear in two ways: first, in participants everyday accounts of planning, where lists and calendars are often used together (Studies 12); and second, where calendar features are embedded within to-do list apps, as in Sunsama in Study 5 or several apps reviewed in Study 4. In these cases, calendars are treated as part of the PTM ecosystem, but the analytic emphasis remains on to-do lists, which more directly expose planning biases and provide clearer design leverage for supporting realistic planning.

1.6 Thesis structure

This thesis contains six chapters. Chapter 1 provides the introduction. Chapter 2 reviews related work. Chapters 3 to 5 present six studies, two for each chapter. Each study is presented in full, with its own discussion section to situate the findings and highlight immediate implications. Finally, Chapter 6 provides a general discussion that integrates findings across all six studies.

Chapter 1 introduces the motivation for the thesis, outlines the research questions and contributions, and explains the significance of accurate task planning in academic contexts. **Chapter 2** reviews relevant literature within HCI, drawing on psychology where it could inform the design of planning technologies, and respective HCI studies that have explore ways to provide PTM support.

Chapter 3 presents the first set of empirical studies. Study 1 investigates the challenges of planning tasks accurately in academic work, identifying widespread overoptimism and the factors that contribute to planning failures in real-world contexts. Study 2 examines how individuals adapted their planning strategies during the Covid-19 lockdown, showing how reflective practices such as

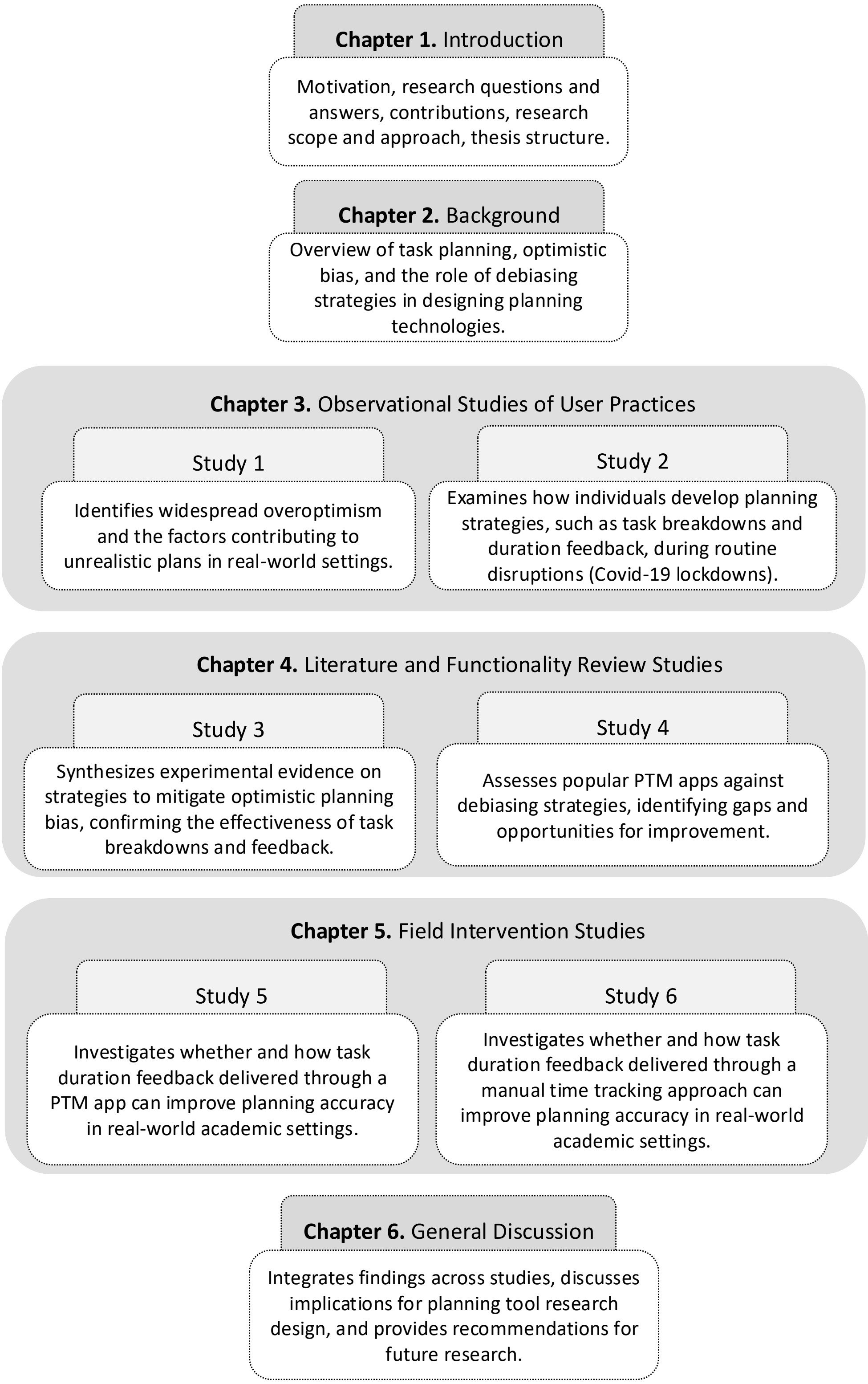


Figure 1.1: Overview of thesis structure

task breakdowns and manual time tracking emerged to support more realistic planning. Together, this chapter establishes an empirical foundation by documenting how planning inaccuracies arise in situ and how everyday strategies can at some cases mitigate them.

Chapter 4 contains two review studies. Study 3 translates findings from psychology by reviewing experimental evidence on debiasing and identifying four strategies relevant to planning accuracy. Study 4 then conducts a functionality review of 47 personal task management (PTM) apps, evaluating the extent to which these strategies are implemented in practice and highlighting significant design gaps. As a whole, this chapter bridges theory and practice, developing a framework that connects psychological evidence with the design of PTM technologies.

Chapter 5 reports two field interventions. Study 5 examines the use of a commercial PTM app (Sunsama) that delivers task duration feedback, while Study 6 tests a lightweight spreadsheet-based tracker. Together these studies evaluate how debiasing interventions operate in real-world academic contexts and how they affect both planning accuracy and users lived experiences. Beyond individual findings, this chapter demonstrates how debiasing strategies can be embedded in everyday tools, highlighting both the promise and the practical challenges of intervention in the wild.

Finally, **Chapter 6** provides a general discussion that integrates findings across all six studies. It highlights the contributions to HCI theory and design, reflects on methodological approaches, considers practical and societal implications, and outlines directions for future research. See Figure 1.1.

Chapter 2

BACKGROUND

Chapter Outline

In this chapter, I review literature on task planning in academic work, the optimistic planning bias, and related de-biasing interventions, establishing a foundation for understanding challenges in planning accuracy and potential avenues for support within academic contexts.

This chapter provides an overview of previous literature to situate my research on task planning in academic work, the optimistic planning bias, and de-biasing interventions. It begins by defining planning within the academic context, outlining its benefits and the specific challenges academics face in balancing multiple responsibilities and projects. Following this, the chapter examines the optimistic planning bias, summarizing experimental and observational studies that document a general tendency toward underestimating task duration. Finally, the chapter explores psychological interventions aimed at addressing this bias and considers the role of HCI research and tools in supporting realistic planning. Together, these sections lay the groundwork for identifying opportunities to enhance the technologies that provide planning support for academics.

2.1 Task planning in academic work

2.1.1 Defining planning

Planning is the process of thinking about the activities required to achieve a desired goal [23]. Defining planning can be challenging because nearly every human action involves some form of planning, even if it occurs only seconds before the action itself. For instance, sitting still or meditating requires a deliberate choice, even if that choice is made moments before the act. At the neurological level, an electrical activity called the *readiness potential* builds up in the brain 1-2 seconds before any conscious action is performed [24]. This reflects an accumulation of evidence and intention even before we are consciously aware of it. Thus, even instinctual or reactive behaviors have a preparatory phase, albeit brief and subconscious.

Rather than contrasting planning with *unplanned* actions, it is more appropriate to view planning as existing on a spectrum. At one end are actions chosen and structured deliberately, involving forethought, goal-setting, and resource allocation to increase the likelihood of achieving desired outcomes. At the other end are minimally planned or reactive actions. Planning, therefore, refers to a conscious, intentional process of deciding on actions in advance, with greater care and strategic decision-making.

Another complexity in defining planning lies in its timescale. Planning can range from approaching a daily task to setting long-term career goals. Longer-term planning is often categorized under *goal-setting* [25], as individuals tend to set broad goals without specific plans for achieving them. This is not to say that creating detailed long-term plans is not considered planning, it is, but the term planning often refers to activities conducted on a regular basis, such as setting out plans for the day or week ahead.

In this thesis, I focus on task planning - deciding what tasks to do and when. Task planning has been previously referred to as short-range planning [4] or time management planning [26]. It ranges from more detailed to more minimal, and it is typically done on a short-range daily timescale. For example, preparing a daily task list in the morning before starting work is wide-spread form of task planning. In the literature, planning is part of the broader domains of time and task management [27, 9], which also encompass other activities aimed at the efficient use of time, such as distraction blocking, delegation of tasks, and collaboration with others.

Lund and Wiese [26] conducted an observational HCI study to understand how individuals engage in task planning in real-world settings. They examined the task planning behaviors among 19 graduate students, focusing on how individuals engage with planning, how tools aid in the process, and whether structured approaches improve effectiveness. They identified two primary aspects of planning: what tasks to perform and when to perform them. Participants demonstrated diverse planning practices, which were categorized into mechanics and methods. Mechanics referred to specific actions like reviewing tasks, setting priorities, scheduling, and estimating time for completion. Participants adapted their approach based on context, such as workload intensity or personal preference. For instance, some created schedules with exact time blocks, while others relied on flexible to-do lists or even left days unplanned depending on their demands (i.e. they had different methods).

Based on their findings, Lund and Wiese define planning as *determining tasks to be performed on a particular day, prioritizing and scheduling the order of such tasks, and sketching out the approximate amount of time to be spent on each task*. Not all these elements need to be present for planning to occur. For example, individuals may set minimal intentions or simply review their calendar, adjusting plans throughout the day. This thesis adopts a similar conceptualization of planning, emphasizing the *what to do* and *when to do* aspects. This type of planning is often closely intertwined with the activity of to-do list making, which serves as a practical mechanism for deciding what to do and when.

To summarize, task planning involves deciding what tasks to perform and when, often using tools like to-do lists. It ranges from detailed schedules to minimal intentions and is a key aspect of time and task management. This thesis adopts Lund and Wiese's definition [26], focusing on prioritizing, scheduling, and estimating time for tasks as part of daily planning practices.

2.1.2 The benefits of planning for productivity

Defining productivity

Planning is widely recognized for supporting productivity. Traditionally, productivity is defined in economic terms as output per worker per unit of time and is a metric often emphasized by employers. However, in knowledge work, productivity is more complex and personal. It includes whether individuals feel they are making progress on tasks they find meaningful, even when those tasks are externally assigned. Guillou et al. [28] challenge narrow definitions of productivity by proposing a more holistic measure: Time Well Spent (TWS). In their study, 40 full-time knowledge workers from diverse occupations participated in a 5-day experience sampling method study. Each day, participants reflected hourly on how they spent their time, how well they felt their time was spent, and how they felt emotionally. At the end of the week, they were asked to define what TWS meant to them.

The findings revealed that participants' definitions of TWS encompassed not just task completion or productivity, but also emotional engagement, a sense of accomplishment, work enjoyment, and self-care. For example, participants reported that time was well spent when they felt in control of their tasks, experienced flow, or took breaks that supported their well-being. Importantly, the study found that reflecting on TWS during the workday led participants to reframe their understanding of productivity to include emotional and holistic dimensions of work. This suggests that productivity for knowledge workers cannot be fully captured by traditional output metrics, and that planning and time management tools should also consider subjective experiences of productivity.

When productivity is defined in these more personal and meaningful terms, where workers feel engaged in valued tasks, experience autonomy, and make progress on goals that matter, it is associated with several positive outcomes. For example, Knight et al. [29] found that when workers perceived themselves as productive in meaningful tasks, they experienced greater well-being and job satisfaction, particularly in environments with autonomy and feedback. Similarly, Kim et al. [30] showed that perceived productivity is shaped by task completion, but also by workers' emotional states, suggesting that productivity is experienced as a holistic phenomenon encompassing both affective and cognitive dimensions. In this thesis, productivity is defined not only as the efficient performance of tasks but also as engaging in meaningful work that contributes to broader positive outcomes, such as job satisfaction and well-being.

How and why does planning improve productivity?

Planning supports productivity through several mechanisms. First, it supports performance by helping individuals follow through on their intentions. A large body of research on implementation intentions, such as, for example, Gollwitzer and Sheeran [31], demonstrates that specifying the when, where, and how of action, typically in the form *If situation Y occurs, then I will do X*, substantially increases the likelihood that goal intentions are enacted. These if-then plans work by enhancing cue accessibility and strengthening the association between situational triggers and responses, thus automating action initiation in critical moments. A meta-analysis of implementation intentions research found that such planning interventions reliably improved goal attainment across a wide range of domains, especially when individuals faced self-regulatory difficulties or competing demands [31]. Recent work by Sefidgar et al. [32] further supports the effectiveness of implementation intentions, showing that when combined with mental contrasting and embedded in a digital planning tool, they significantly improve plan adherence..

Research based on self-regulation theory [33] explains that planning helps people monitor their progress, get feedback, and notice gaps between where they are and where they want to be. This process activates motivational resources and workers are more likely to invest effort when they can clearly see what needs to be done and how far they've come [34]. Rogers et al. [35] extended this evidence by demonstrating that even simple prompts to make concrete plans significantly improve follow-through on real-world behaviors, such as voting or attending health appointments. These planning prompts are effective because they help individuals anticipate obstacles, structure their day around goal-relevant actions, and reduce reliance on willpower at the moment of decision-making. Austin and Vancouver [36] argue that well-defined goals, supported by effective plans, improve focus and resource allocation by narrowing attention and reducing decision fatigue.

Planning also frees up cognitive resources. Masicampo and Baumeister [37] showed that forming concrete plans reduces intrusive thoughts about unfinished goals, allowing individuals to engage more fully in their current tasks. By reducing the need to repeatedly decide what to do next, planning supports cognitive clarity and reduces the burden of multitasking and context switching. This is especially important in digital knowledge work, where constant interruptions and distractions undermine attentional control and productivity [38, 12]. In addition, Syrek et al. [5] demonstrated through a 12-week diary study that, within individuals, higher levels of unfinished tasks at the end of the workweek triggered increased affective rumination over the weekend, which in turn impaired sleep quality.

Finally, a less examined but increasingly acknowledged benefit of planning lies in its ability to prompt reflection and intentionality. Rather than serving solely as a means of execution, planning offers structured opportunities to evaluate goals, revise intentions, and recognize misalignments between current behaviors and personal values. In a scoping review of personal informatics tools, Ekhtiar et al. [39] identify reflection as a core design implication for effective goal setting. Their analysis shows that well-designed goal-setting tools often incorporate reflective features to help users better understand their motivations, constraints, and evolving priorities. Similarly, Agapie et al. [40] demonstrate in a longitudinal study that the process of planning, particularly when situated in weekly reviews, elicits self-evaluation and strategy adjustment. They find that this reflective engagement is essential to sustaining planning over time. Rooksby et al. [41], in their study of *lived informatics*, further emphasize that people use self-tracking tools, including some planning tools like to-do lists, not only to carry out predefined goals but to explore what matters to them through ongoing use. Across these studies, planning emerges as a practice that supports self-understanding as much as task completion. This reflective aspect is especially valuable in knowledge work, where goals are often ill-defined and evolve over time. By prompting users to reconsider what they are pursuing and why, planning can enable more intentional action, ultimately fostering productivity that is personally meaningful rather than narrowly output-driven.

Taken together, planning contributes to productivity not only by increasing task completion but by shaping how individuals experience their work. It helps translate intentions into actions, supports attentional focus, and supports a sense of agency. These mechanisms are especially relevant in knowledge-intensive and autonomous work contexts, where the ability to self-direct and reflect is fundamental to both performance and well-being.

2.1.3 The nature of academic work for early career researchers

Early Career Researchers (ECRs) are defined in this thesis, following UKRI and Wellcome guidance, as individuals from the start of doctoral training up to approximately eight years post-PhD, excluding formal career breaks. ECRs occupy a career stage characterised by high responsibility, developing autonomy, and considerable structural constraint. They must balance research, teaching, administration, and career development, each with different temporal demands. Long-term research requires sustained, self-directed work, whereas teaching and institutional duties impose fixed schedules. This combination creates planning challenges in which flexibility and external constraint are in constant tension.

Evidence shows that workloads in academia have intensified over recent decades, with rising publication expectations, competitive funding environments, and expanding administrative responsibilities [42, 43, 44]. These pressures fall disproportionately on ECRs, who report high levels of quantitative overload and frequently work beyond standard hours [45, 46]. Such overwork is closely associated with stress, worklife conflict, and poor mental health outcomes [47, 48].

The nature of academic tasks contributes to this strain. ECRs manage multiple and often conflicting roles, and role overload is consistently identified as a major stressor [49]. Studies document increasing administrative burdens in UK higher education [45], which ECRs often shoulder without the support structures available at later career stages [50]. Role conflict and overload have been linked to emotional exhaustion and attrition from academic careers [44, 51].

Self-management is central to ECR work. Historically, autonomy buffered academics from stress, but recent evidence shows that autonomy has become more burdensome under conditions of job insecurity and performance measurement [11, 52]. ECRs are expected to prioritise complex, fragmented workloads despite having limited control over structural constraints. Cultural norms within academia further intensify these pressures. Overwork is widely normalised: late-evening work, weekend working, and blurred boundaries are viewed as routine rather than exceptional [18]. ECRs often internalise these expectations and respond by overcommitting [53].

2.1.4 Planning practices and gaps in HCI research

Understanding planning practices and tool use

A foundational line of HCI research has focused on understanding how individuals plan and manage tasks in everyday settings. Rather than assuming a single ideal tool, studies have revealed the diversity of strategies people adopt, how tool choices evolve over time, and the practical factors that shape planning behaviors.

One of the earliest and most influential studies in this space was conducted by Blandford and Green [9]. Departing from efforts to build all-in-one digital planners such as Meeting Maker, they examined real-world assemblages of planning tools among knowledge workers. Through interviews with 16 staff members, including managers, administrators, and academics, they found that participants relied on a variety of tools in parallel: paper diaries, post-it notes, personal digital assistants (PDAs), physical object cues (like piles of documents), and Meeting Maker itself. Importantly, 75% still relied on paper diaries, and 56% used Meeting Maker, often adapting general-purpose tools like email to support planning. This plurality was not seen as a flaw but a necessity: no single system could support all needs.

They identified several factors that shaped tool choice, including portability, accessibility, shareability, and ease of updating. Participants often encountered breakdowns in planning practices during busy periods, leading to a shift from structured systems to more informal strategies. Rather than advocating for a singular *ideal* tool, Blandford and Green concluded that integration across tools was key to supporting the dynamic and distributed nature of real-world planning.

Subsequent work has reinforced and expanded these insights. Hu et al. [6], for example, combined a formative study (interviews and questionnaires) with a large-scale survey (*N*=248) to investigate planning tool preferences among knowledge workers. They identified six latent dimensions shaping tool choice: Communicability, Structure, Portability, Adaptability, Physicality, and Visualizability. These dimensions reflected both individual personality traits and work context factors.

Similarly, Haraty [54, 55] explored how personal task management behaviors evolve over time. Their studies categorized users as Adopters (structured digital tool users), DIYers (customizers of general-purpose tools), and Make-Doers (minimalists relying on basic systems). They found that users frequently shifted planning strategies, adapted tools, or changed their toolsets altogether in response to changing goals, dissatisfaction, or new opportunities. Rather than viewing planning practices as fixed, these studies emphasized the importance of flexibility and support for transitions.

In addition, several studies have emphasized that planning tools often fail under stress. Through interviews and video diaries with 26 university staff, Kamsin et al. [2] and Bellotti et al. [1] found that PTM systems frequently collapsed during peak workload periods not because users lacked prioritization skills, but because tools failed to account for internal factors such as energy, motivation, and emotional state. Participants described abandoning digital systems during periods of high stress in favor of ad hoc strategies that felt more manageable. These findings argue for planning tools that dynamically adapt to users' real-time cognitive and emotional capacities.

Similarly, in [1], Bellotti and colleagues conducted a four-week observational study tracking everyday task management practices. They found that participants often demonstrated sophisticated prioritization strategies using calendar blocking, lightweight state tracking, and value-based grouping, contrary to the stereotype of poor prioritization. Prioritisation here refers to the set of strategies participants used to determine what to focus on next, how to sequence tasks, and how to allocate time and attention based on importance, urgency, personal values, and current mental or situational constraints. High task completion rates (81%) were reported. Failures were typically caused by external disruptions (e.g., unexpected meetings). The study called for tools that support evolving task representations rather than imposing rigid structures.

Toward supporting effective planning outcomes

While this body of work provides rich insights into planning behaviors and breakdowns, much less attention has been paid to whether current practices and tools actually lead to effective outcomes. In this thesis, the reported studies aim to shift focus from how people plan to how well they plan, aiming to identify design principles that better support realistic planning and task completion.

Existing research richly describes what tools people use and why they prefer them, but it rarely interrogates the quality of the resulting plans, such as whether users accurately estimate task duration, appropriately scope tasks, or realistically manage competing demands. By focusing on planning effectiveness, the research in this thesis can complement prior studies and move toward tools that not only support users' current preferences but also help them improve their planning strategies over time.

This perspective opens new opportunities for intervention. Instead of simply designing more flexible or integrated planning tools, mechanisms for detecting when planning is likely to fail can be explored, such as identifying patterns of over-optimistic time estimates or recognizing when fragmentation is undermining deep work, and providing timely feedback. In doing so, this thesis aims to contribute to the development of tools that support not just task management, but sustainable, effective self-regulation in complex work environments.

The next section turns to existing research on planning accuracy. While the first empirical study of this thesis began from a broad exploration of why academic plans so often go awry, one explanation proved especially relevant: the optimistic planning bias (See Chapter 1, Study 1). To keep the structure coherent, the optimistic planning bias is introduced here in the related work, even though it only emerged as a focal point after Study 1. This positioning ensures that subsequent chapters can situate the empirical findings within both the psychological literature on bias and the HCI literature on planning tools. This sequencing is consistent with HCI research practice, where initial empirical inquiry is used to surface the most salient constructs, which are then situated within existing theoretical frameworks [56].

2.2 The optimistic planning bias

2.2.1 Definition and theory

The optimistic planning bias refers to the tendency of individuals to underestimate the time required to complete tasks, even when they have past experience indicating that these tasks may take longer. This bias, also called the *planning fallacy*, introduced by Tversky and Kahneman [57], describes how people's predictions are overly optimistic due to a reliance on best-case scenarios and neglect of historical evidence. The Sydney Opera House is often cited as the classic example for this bias. The estimated project timeline for the construction was 4 years with budget of 4 million dollars however it took 14 years and it ended up costing 100 million dollars.

At its core, the optimistic planning bias arises from individuals focusing on the internal details of a task - what needs to be done, how it should proceed, and under ideal conditions - while neglecting external factors such as unforeseen interruptions, delays, or competing demands. Although in principle the bias encompasses both underestimation and overestimation, research typically foregrounds underestimation because it occurs more frequently and has greater practical consequences. The distributional vs. singular information theory [58] provides a foundational explanation for the optimistic planning bias. According to this theory, individuals often neglect distributional (general or population-based) information, which provides an aggregated sense of how tasks typically unfold, and instead rely on singular information based on their unique perspective or experience of a specific task. This preference for singular information often results in an over-reliance on best-case scenarios, ignoring the variability and potential disruptions that might occur during task execution.

Memory bias is another contributing factor [59]. Previous task durations are frequently misremembered, often being recalled as shorter than they were. This leads to a distorted basis for future predictions, reinforcing overly optimistic planning tendencies. For instance, individuals tend to view past delays as isolated or exceptional events, underestimating the likelihood of encountering similar setbacks in the future.

To summarize, the optimistic planning bias, or planning fallacy, describes individuals' tendency to believe they can achieve more in less time due to a focus on best-case scenarios and neglect of past evidence. This bias arises from relying on singular task details rather than distributional information that accounts for variability and potential disruptions. Memory bias further exacerbates the issue, as people often misremember task durations as shorter, leading to repeated overly optimistic predictions.

2.2.2 Experimental studies

The optimistic planning bias has been demonstrated in numerous experimental studies across various domains, consistently revealing that individuals underestimate the time required to complete tasks, even when they have prior knowledge or relevant experience. These studies highlight the systematic nature of the bias and its persistence across diverse contexts.

A review by Halkjelsvik and Jørgensen [60] provides evidence of the optimistic planning bias across tasks ranging from simple manual activities, such as origami construction, to more complex, knowledge-intensive activities, such as software development. For example, in origami tasks, individuals predicted significantly shorter completion times than observed, despite having prior exposure to similar tasks. Their analysis found that while longer tasks were frequently underestimated, shorter tasks were more likely to be overestimated. This effect can be explained by cognitive heuristics: for shorter tasks, individuals may overcompensate for potential delays, resulting in conservative overestimations, while for longer tasks, optimism about smooth execution and neglect of cumulative delays leads to underestimation. This pattern suggests a general tendency to focus on idealized task execution while neglecting potential challenges.

In another seminal experiment by Buehler, Griffin, and Ross [61], participants were asked to predict how long it would take to complete academic assignments, such as writing essays. Despite being reminded of their previous experiences with similar tasks, participants' estimates were overly optimistic. On average, their predictions were closer to best-case scenarios than to the actual completion times. Importantly, even when participants were explicitly instructed to base their predictions on past experiences, they continued to make optimistic judgments, reflecting a deep-rooted bias in time estimation.

Further evidence of the bias comes from experiments in software development, as reviewed by Halkjelsvik and Jørgensen. In one study [62], professional software developers were asked to estimate the time required to complete coding tasks. Despite their expertise and access to historical data, developers consistently underestimated task durations. This underestimation was attributed to their reliance on the internal perspective, focusing on the steps required for task completion rather than considering external factors such as interruptions or unforeseen complexities.

A related experiment by Roy et al. [59] demonstrated the bias in group tasks. Participants working collaboratively on project planning tasks systematically underestimated the time required, even when encouraged to discuss potential obstacles as a group. The results showed that group discussion often amplified optimism, as team members reinforced each others' belief in smooth task execution, leading to further underestimation of completion times.

Experimental studies have also highlighted the influence of task complexity on the magnitude of the bias. For simpler tasks, such as solving math problems or assembling furniture, individuals exhibited moderate optimism. However, for more complex tasks involving multiple stages or dependencies, such as project management or academic writing, the bias was significantly larger [61]. These findings suggest that task complexity exacerbates the optimistic planning bias, as individuals are less able to anticipate the cumulative effects of delays and interdependencies in complex tasks.

To summarize, experimental studies consistently demonstrate the optimistic planning bias across diverse domains, showing a persistent underestimation of task durations even when individuals have prior experience. This bias is amplified in complex tasks due to difficulty anticipating delays and interdependencies. Research highlights that both individual and group estimates are overly optimistic, with groups often reinforcing idealized scenarios. These findings underscore the systematic nature of the bias and its resistance to correction, even when individuals are reminded to consider past experiences or potential obstacles.

2.2.3 Observational studies

Observational studies have shed light on how individuals systematically underestimate the time required for task completion in real-world settings. These studies provide insights into the manifestations of the optimistic planning bias across various task types and contexts and point to its pervasive nature. However, they also reveal limitations in understanding the exact magnitude and nuances of the bias across diverse work environments.

Newman [15] demonstrated that busy professionals often fail to complete their daily to-do lists due to overly optimistic expectations of how quickly tasks can be completed. This phenomenon reflects the broader planning fallacy, where individuals maintain the belief that a task will proceed as planned, despite evidence from past experiences suggesting otherwise. Newman's diary-based study asked participants to create daily plans and record actual task durations. The results showed that information work tasks, such as writing and analysis, were consistently underestimated, while time estimates for face-to-face meetings were often more accurate. Furthermore, the study suggested that social tasks, like meetings, were sometimes completed faster than planned to compensate for the slower-than-expected execution of information work.

Claessens et al. [3] extended this line of inquiry by focusing on the prioritization of tasks in daily planning. In their study, research and development employees tracked their planned and completed tasks over three weeks. Participants did not complete 27% of their daily planned work, with urgent but less important tasks often taking precedence over more significant, long-term goals. This prioritization of urgency over importance disrupted their ability to complete planned work effectively. However, while this study highlighted the consequences of task prioritization on planning accuracy, it lacked a detailed task-level analysis of time estimation bias, leaving gaps in understanding the specific factors influencing optimistic planning.

Other studies, such as Bardram and Hansen [63], observed similar tendencies in high-stakes environments like hospitals, where less than half of planned procedures adhered to their original time estimates. This study attributed the optimistic bias to external disruptions and competing priorities but did not provide granular insights into how specific types of tasks contributed to the observed deviations. Similarly, while Bellotti et al. [1] documented delays and task overruns in corporate settings, their findings focused more on interruptions and re-prioritizations than on systematically quantifying the degree of underestimation across different task categories.

The observed differences in estimation accuracy across task types, such as meetings versus information work, point to the importance of examining the optimistic planning bias in greater depth. While these studies provide important evidence of the optimistic planning bias, they often lack the precision needed to quantify the exact magnitude of the bias across various task types and contexts. For example, the relationship between task length and bias, where shorter tasks may be overestimated and longer tasks underestimated, has been noted but not systematically explored. Moreover, many of these observational studies rely on general task categories, without distinguishing between the nuances of specific knowledge work activities, such as email management, data analysis, or collaborative projects.

This limited understanding of the magnitude and variability of the optimistic planning bias highlights the need for further observational research. There is a need for studies that aim to provide task-level analyses of time estimation biases across diverse work contexts, ultimately informing the design of tools and systems that support more accurate and reliable planning.

To summarize, observational studies reveal the pervasive nature of the optimistic planning bias in real-world settings, highlighting consistent underestimation of task durations across diverse contexts. While tasks like meetings are often estimated more accurately due to their structured nature, open-ended tasks such as writing or analysis are particularly prone to the bias. These studies point to the need for further task-level analyses to better quantify the magnitude and variability of the bias and to inform the development of tools that support more accurate and effective planning strategies.

2.2.4 Debiasing methods

Popular planning frameworks

Popular planning frameworks have long been promoted as tools to enhance productivity and time management, yet they often fail to address the root issue of creating realistic plans. Instead, these frameworks typically focus on helping individuals cope with the consequences of overloaded tasks or improving their ability to follow through on existing intentions. This approach is problematic, as it overlooks the importance of learning how to make realistic plans in the first place, which could prevent task overload and procrastination more effectively.

For example, Stephen Covey's First Things First framework [64] emphasizes prioritizing tasks based on urgency and importance, using the Eisenhower Decision Matrix [65]. While it provides a useful structure to manage immediate and long-term priorities, it does not offer detailed guidance on how to plan for multiple urgent tasks or how to avoid underestimating time requirements for important but non-urgent activities. The framework addresses the common tendency to prioritize urgency over importance, but it stops short of teaching users how to create realistic and achievable plans under complex or competing demands.

David Allen's Getting Things Done (GTD) method offers a step-by-step approach to organizing tasks through processes like collecting, processing, and categorizing them into actionable items [66]. While GTD provides specific workflows for managing tasks, such as deferring or delegating them, its primary focus is on organizing and executing tasks rather than addressing cognitive biases like optimism in planning. Additionally, despite the popularity of GTD tools, such as task management apps, there is little empirical evidence to suggest that these tools help users improve the realism of their plans.

The SMART Goals framework developed by Locke and Latham is another widely recognized approach, emphasizing that goals should be Specific, Measurable, Achievable, Realistic, and Timely [25, 67]. While research supports the motivational benefits of clear and challenging goals, the framework does not address how to manage competing priorities or create accurate time estimates for complex, multi-stage tasks. As a result, it focuses on optimizing goal-setting rather than equipping users with the skills to create realistic and adaptive plans.

All three frameworks aim to improve planning and productivity by encouraging prioritization and goal clarity. However, they do not adequately address the challenges of realistic planning in modern knowledge work, where individuals frequently face competing demands, fragmented workflows, and the cognitive biases that contribute to overly optimistic plans. Moreover, none of these frameworks provide tailored advice for specific user groups, such as knowledge workers in different industries or at varying career stages. Academic professionals, for instance, may need planning strategies that address long-term research goals alongside short-term teaching and administrative responsibilities. Similarly, frameworks lack integration with the tools workers already use, such as calendars, email systems, or project management software, limiting their practical application.

Experimental factors

The optimistic planning bias has been extensively studied in psychology through controlled experiments, leading to the development of various strategies aimed at reducing this bias. However, these methods often produce mixed results and are rarely designed with practical implementation in mind. Bridging the gap between psychological research and real-world applications remains a significant challenge for developing planning tools that effectively support users in creating more realistic plans.

One commonly studied strategy is task decomposition, which involves breaking tasks into smaller, manageable components. Connolly and Dean [68] found that this approach encourages individuals to consider the time required for each sub-task, resulting in more realistic time estimates. However, its effectiveness decreases for tasks that are poorly understood or overly complex, where individuals may still overlook critical steps. This limits the practical applicability of task decomposition in real-world scenarios where complexity and uncertainty are common.

Another approach, unpacking tasks by listing all the necessary steps and potential obstacles, has shown promise in encouraging individuals to think beyond idealized execution scenarios. Kruger and Evans [69] demonstrated that unpacking can reduce underestimation by making people more aware of the potential challenges they may face. Similarly, generating pessimistic scenarios, as suggested by [70], improves planning accuracy by prompting individuals to consider potential delays. However, these methods often require substantial cognitive effort, which can make them impractical for everyday use, particularly when dealing with complex tasks under time constraints.

Providing feedback on past time estimation errors has been shown to consistently improve prediction accuracy. Tobin and Grondin [71] demonstrated that repeated feedback cycles help individuals recalibrate their expectations and align their estimates more closely with actual task durations. Further, framing estimation questions differently has also been found to influence the accuracy of predictions. Griffin and Buehler [72] showed that emphasizing challenges or difficulties in task execution, rather than simplicity, leads to more conservative and realistic estimates. Perspective-taking techniques, such as imagining how a peer might estimate the same task, can also reduce bias. Buehler et al. [61] found that individuals tend to make more accurate predictions for others than for themselves, and adopting this external perspective can help mitigate overly optimistic judgments. However, these methods are not naturally embedded into existing planning practices or tools, limiting their utility in real-world settings.

Despite these results, the effectiveness of these debiasing methods remains inconsistent. For example, while strategies like feedback and task decomposition show potential, others, such as surprise listing, where individuals are encouraged to anticipate unexpected delays, yield mixed results. Moreover, these strategies often require controlled experimental environments or impose high cognitive demands, which hinder their practical adoption in dynamic and fast-paced work settings.

The insights gained from these psychology experiments primarily focus on understanding cognitive mechanisms rather than providing actionable solutions. This creates a gap between theoretical approaches and practical applications. Effective planning tools must go beyond usability and integrate debiasing strategies in ways that are accessible, intuitive, and compatible with users' existing workflows. Features such as feedback on task durations, importance-based prioritization, and prompts for scenario planning could help bridge this gap by making these methods more actionable for knowledge workers. This thesis aims to suggest concrete evidence-based ways to design such features in human-centered ways.

To summarize, psychological research provides a strong foundation for understanding the optimistic planning bias and offers several promising strategies for addressing it. However, these methods often remain theoretical and require translation into practical tools to be effective in real-world contexts. This thesis seeks to bridge this gap by designing and evaluating HCI interventions that integrate these insights to improve planning accuracy and support users in managing their tasks more effectively.

HCI debiasing studies

HCI research on addressing time estimation biases is still in its early stages, with limited studies directly targeting optimistic planning bias. While much work has been done on tools that enhance productivity, focus, or task execution, the challenge of improving planning accuracy has received less attention.

A noteworthy study by Koval et al. [73] explored the use of visualizations to mitigate optimistic planning bias. Participants used a tool that provided feedback on past time estimates and actual task durations. By comparing these measures, participants were able to recalibrate their expectations for future planning. The results showed a reduction in planning bias and suggested that visualization techniques hold promise for helping users align their planning more closely with reality. However, the study noted that such interventions require careful design to avoid overwhelming users with data and to ensure usability in everyday contexts.

Another relevant investigation, Zhou et al. [74], examined the impact of retrospective estimations on stress and perceived control in social network usage. Using the RescueTime application, participants were asked to estimate their daily usage of social networking sites while being provided with objective measures of their behavior. While the study did not directly address planning, it found that participants' estimation accuracy improved over time, leading to reduced stress and an increased sense of control. This highlights the potential of combining estimation tasks with feedback mechanisms to influence user behavior positively.

In a related auto-ethnographic study, Cox [75] used the HoursTracker app to track her work hours. Initially, she discovered discrepancies between her perceived and actual work hours, leading to a "digital epiphany" about her workload. This reflection prompted adjustments in her planning and work habits, reducing stress and improving task management. While the study was self-reported, it underscores the potential of manual tracking for fostering awareness and encouraging planning improvements.

Despite these insights, most HCI studies have focused on tools for managing distractions and improving task execution rather than addressing the core issue of planning accuracy. For instance, tools like TimeAware and TimeToFocus aim to help users minimize time spent on interruptions but do not engage with the cognitive biases that lead to unrealistic planning. These tools often prioritize enhancing productivity rather than fostering better planning strategies.

To summarize, while existing HCI research provides valuable insights into feedback and reflection, it has largely overlooked the specific challenges of optimistic planning bias. More work is needed to translate psychological theories into actionable, user-friendly tools that enable individuals to make more realistic plans, recalibrate expectations, and anticipate potential disruptions effectively. This represents a critical gap and an opportunity for future HCI research to develop tools that directly address planning accuracy within diverse real-world contexts.

2.2.5 Summary of Chapter 2

This chapter establishes the conceptual and empirical foundation for the thesis by reviewing literature on task planning in academic work, the optimistic planning bias, and de-biasing interventions. It begins by defining task planning as the process of deciding what to do and when, grounded in research from time management and HCI. Task planning is shown to support not only task completion but also focus, cognitive clarity, and reflectionkey aspects of productivity, especially in knowledge work. The chapter then examines the challenges of academic work, including high workload, role conflict, cultural norms of overwork, and increasing task fragmentation, all of which make effective planning more difficult but also more essential. Planning becomes a cognitive scaffold that supports both task execution and well-being. The chapter next reviews HCI research on planning practices, highlighting the diversity of tools and strategies people use, and how existing planning systems often break down under stress or fail to support deeper planning needs. It identifies a gap in HCI literature: while many studies explore how people plan, few assess how well they plan or examine outcomes such as accuracy or plan adherence.

The second half of the chapter explores the optimistic planning biaspeople's tendency to underestimate how long tasks will take. Drawing from psychology, it details theoretical explanations (e.g., reliance on best-case scenarios, memory biases) and experimental and observational evidence showing the bias is persistent across domains, especially for complex or open-ended tasks like writing or analysis. Finally, the chapter reviews debiasing strategies from psychology (e.g., task decomposition, feedback, scenario planning), popular productivity frameworks, and early HCI interventions. While promising, most existing strategies are cognitively demanding or poorly integrated into tools, highlighting the need for user-friendly systems that support more realistic planning in practice. The chapter concludes by identifying a gap: few tools support users in creating accurate, feasible plans. This motivates the thesis's focus on designing and evaluating interventions that help academics plan more realistically and manage their complex workloads more effectively.

Chapter 3

Studies 1 and 2: Understanding existing planning practices

Parts of this chapter have been published in Ahmetoglu, Brumby and Cox [76, 77, 78].

Chapter Outline

In this chapter, I present the findings from two studies that examine the challenges of accurate task planning in academic knowledge work. Study 1 uses a mixed-methods diary and interview design to investigate when, how, and why planning is inaccurate in everyday work. Study 2 is a longitudinal qualitative study that explores how planning routines were disrupted and rebuilt during the COVID-19 pandemic. Together, these studies provide empirical insights into the roots of planning inaccuracies, the role of disruption in shaping planning behaviour, and the strategies that workers adopt to navigate these challenges.

This chapter addresses the first research question of the thesis: *What are the main challenges in accurate task planning at work?* Study 1 begins with a quantitative analysis of daily planning errors, revealing systematic discrepancies between planned and actual work, that align with an optimistic planning bias. It then explores the contextual and cognitive factors contributing to these discrepancies, identifying mechanisms such as omitting preparatory work and breaks, responding to unexpected requests, and experiencing fatigue. The analysis also surfaces four distinct planning strategies (minimal, daily, weekly, and multi-level) that varied in frequency, detail, and time-estimation needs, highlighting the diversity of planning behaviours that future tools should support.

Study 2 builds on these findings by examining how such challenges evolve under conditions of disruption. Drawing on 76 weekly interviews conducted during the first COVID-19 lockdown, it shows how academic workers' planning routines broke down and were gradually rebuilt. The findings highlight widespread disengagement from planning, emphasising the importance of systems that support not only task organisation but also motivation and re-engagement. They also reveal two concrete strategies (task breakdown and manual time tracking) that helped participants regain control and develop more accurate and sustainable habits. These strategies, adopted proactively by participants over the course of the study, suggest promising directions for design.

The chapter concludes by arguing that supporting planning accuracy is both feasible and beneficial, particularly when approached through strategies that can be integrated into daily practices. This sets the stage for subsequent studies that investigate how such strategies can be operationalised and supported by digital planning technologies.

3.1 Introduction to Study 1

This chapter addresses the first research question of the thesis: *What are the main challenges in accurate task planning at work?* As a foundational step toward designing more effective planning technologies, it aims to build a grounded understanding of when, why, and how planning breaks down in everyday knowledge work. While later chapters evaluate existing tools and test interventions for reducing planning bias, this study provides the empirical starting point by closely examining the nature and consequences of task planning inaccuracies in situ.

Prior research in HCI has documented that individuals frequently fail to complete a substantial portion of their planned work. For instance, Bellotti et al. [1] found that senior employees underestimated task durations in 40% of cases, resulting in stress and difficulty meeting priorities. Similarly, Newman's diary study showed that optimistic daily plans led knowledge workers to abandon important tasks and sacrifice personal time to catch up [15]. Claessens et al. [3] reported that 27% of planned tasks remained unfinished, with workers often reacting to short-term urgencies rather than making progress on meaningful goals. These breakdowns in task planning can contribute to chronic overload, work-life conflict, and burnout.

While these studies point to persistent issues in planning, most focus on identifying design requirements for digital task management tools rather than addressing the root causes of planning inaccuracies. The resulting design recommendations often assume that planning errors are inevitable and propose features to work around them, such as avoiding reliance on time estimates, rather than helping users plan more accurately in the first place. For example, Bellotti et al. [1] suggested designing systems that sidestep precise time predictions, while Newman's [15] augmented diary study captured time estimation errors but focused primarily on writing tasks. This leaves unanswered questions about which types of tasks are most prone to bias, what users perceive as the causes of delays, and which personal planning strategies may support more accurate outcomes.

This chapter addresses these questions by expanding on the augmented diary methodology to include a broader range of task types and by incorporating follow-up interviews to contextualize planning outcomes [15]. The study explores the following three specific research questions:

1. **The How of planning inaccuracies**: Are certain types of tasks more prone to inaccurate planning than others?

2. **The Why of planning inaccuracies**: What are the main reasons for delays in planned daily work?

3. **The When of planning inaccuracies**: How effective are the different strategies that workers use for task planning in creating accurate plans?

This study contributes a first step toward rethinking planning support tools, not as digital replicas of analog systems, but as intelligent aids that can adapt to different contexts, account for common planning biases, and help individuals plan in ways that are both more accurate and psychologically sustainable.

3.2 Method of Study 1

3.2.1 Participants

Twenty participants (nine male) took part in the study with a mean age of 29 years (*SD* = 4.8 years). They were academics and early career researchers at UK universities (3 x lecturers, 2 x postdoctoral researchers, 13 x PhD students and 2 x internship graduate students). They were recruited through a departmental call to participate distributed in social channels. Participation was voluntary and participants were able to drop out at any time. The study was approved by the UCL Ethics Committee.

3.2.2 Design

The study used a mixed-methods approach consisting of one day augmented diary (morning plan and a diary), and two semi-structured interviews: one as a follow up (Interviews 1) to the diary and one that explored participants' views of their own planning (Interviews 2). All participants filled in a two stage report: they indicated their plans in the morning and reported their behaviours continuously in a diary throughout the day. After, they took part in Interviews 1 on the same or next day and another structured interview Interviews 2 during the same or following week.

3.2.3 Procedure

Participants listed in the morning all tasks they aimed to achieve during their workday and were asked to estimate the likely duration for each task. They used pen and paper report forms. They were free to consult their calendars or to do lists to remind themselves of their agendas. The morning list was then handed to the researcher and was not given back to the participants until the end of the workday. After filling the morning lists, participants were asked to continue their workday as usual while keeping a pen and paper record of the main task they decided to work on. They had to report the start and end times of each task with as many details as possible.

After the diary day was over, plans and diaries were inspected side by side with the participant in Interviews 1 taking place either immediately or a day after the diary observation. Participants were able to elaborate on the discrepancies between their plans and actual activities and reflect on the reasons to why some tasks were completed but not planned, planned but not completed, or executed in a different way than originally planned. During Interviews 1, it was emphasised that the study aimed to understand the circumstances that led to changes in planned activities and that those changes were completely normal and expected as opposed to faults on the part of the participant. Additional details about the tasks were also gained. For instance, there was discussion on whether the discrepancy was surprising to the participants and if it happened often, and which (if any) discrepancies the participant found as challenge compared to "part of the job" change of plans.

Can you explain what *planning work* means at your job?
Do you plan your work? Why/Why not?
How do you feel about planning? Does it work for you? Why not?
How do you decide which tasks to start your day with? Think about for a moment. Why?
How do you plan tasks that you have no prior experience with?
What tools do you use to plan; at what times; at what devices; and why do you choose them?
Are you aware of the factors that make you feel productive while working?
If you could imagine any tool, what would be the perfect tool to boost your productivity?

Figure 3.1: Interview 2 Guide used in Study 1

In addition to the diary method, participants also took part in semi-structured interviews which were aimed at exploring the daily planning strategies and tools participants used. The interview guide for Interviews 2 is shown in Figure 3.1 Interviews lasted on average 30 minutes, and they took place either in person at a university office or by telephone. Interviews were recorded by the researcher using notes in Microsoft Word.

3.2.4 Data analysis

Plans and diaries were transferred to digital tables (see Figure 3.2). Each reported activity was anonymized by deleting information about specific projects and people. Lunch time was omitted from analysis to ensure that participants felt as minimally observed as possible. A new comparison table was created for each participant with estimated time and actual time spent on all reported tasks. This table was used to compare the estimated and actual total workday duration for each participant. The data met the assumptions of normality required for t-tests.

All tasks were then thematically analysed and sorted into different categories, for instance, writing research, scheduled meetings and coding. The accuracy of estimates for different categories of tasks was analysed (see Figure 3.2). In cases where a participant had more than one task of a given type (e.g. to read two separate works), the average duration was used for mean statistics for each type of task (e.g. reading research). Finally, a 20% threshold was used to sort categories of tasks into correctly predicted (less than 20% bias), underestimated (took 20% longer) or underestimated (took 20% shorter) (following prior work, e.g. [79]).

A

What do you aim to achieve during the day?	*For how long?*
Work on research report	3 hours
Supervisory meeting	13:30-14:30
Transcribe audio data	2 hours
Reply to emails and get updates	1 hour
Help team with coding project	1 hour

B

Describe your current task	*Start*	*End*
Email and communications	9:40	10:37
Writing report	10:37	11:08
Email and communications	11:08	11:16
Socialising	11:16	11:26
Email and communications	11:26	11:45
(Lunch break)		
Writing report	12:21	13:20
Meeting	13:30	14:30
Fixing coding issues	14:35	17:00

C

Type of task	*Estimated time*	*Spent time*
Research report	180 min	89 min
Meeting	60 min	60 min
Transcribe	120 min	0 min
Email and communications	60 min	79 min
Coding task	60 min	145 min

Figure 3.2: A: Morning list with planned tasks and estimated duration. B: Diary form with reported tasks, start and end times. C: Comparison table used for analysis showing estimated and spent time on each task.

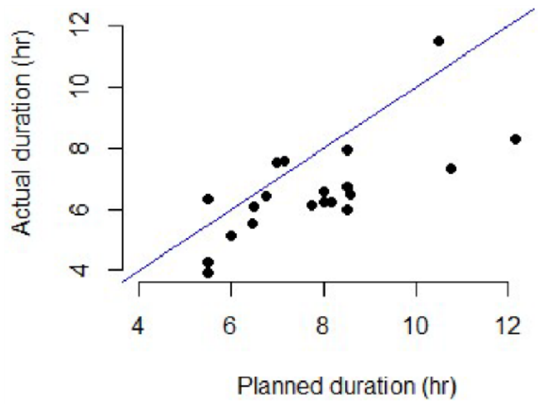
3.3 Results of Study 1

Results are presented in three sections. First, a quantitative analysis of the diaries is presented. This gives insights about the proportion and nature of work tasks which were executed as planned and those which were not executed as planned. Second, a qualitative analysis of Interview 1 is presented. This focuses on participants' explanations for changing their plans later in the day. Third, a qualitative analysis of Interview 2 is presented. This focuses on exploring people's strategies to execute their work in a timely manner.

3.3.1 Accuracy of time estimates in daily plans

Time estimation bias in workday duration

Figure 3.3 shows the planned (estimated) work duration and actual workday duration for each participant. The average duration of workday tasks was estimated to be 7 hrs 44 min (SD = 102 min) whereas the actual duration of workday tasks reported was 6 hrs 40 min (SD = 101 min), including breaks and unplanned tasks but excluding lunch time. A paired sample t-test suggested that participants planned to work for significantly longer than they actually did, t (19) = 4.01, p = .001.



Type of tasks	*N*	*Estimated*	*Actual*
Correct			
Scheduled meetings	16	78 (43)	64 (42)
Scheduled seminars	6	130 (113)	110 (95)
Reading research	11	68 (56)	66 (31)
Creating presentation	5	91 (27)	77 (52)
Underestimation bias			
Email and communications	12	51 (35)	68 (55)
Coding and data analysis	5	90 (30)	182 (240)
Overestimation bias			
Writing research	12	119 (78)	78 (54)
Planning and decision-making	5	63 (50)	19 (28)

Figure 3.3: On the left: The association between planned (estimated) work duration and actual workday duration for each participant. The blue line shows a perfect relationship with no bias. Data points below the line indicate that the planned duration was longer than the actual one. On the right: Summary of time estimated and time spent on different types of tasks. N = number of participants. Time values are $Mean(SD)$ in minutes

Proportion of planned and unplanned work done

Out of the reported 6 hrs 40 min spent on work tasks during the day, 54 min (SD = 50 min) were spent on work activities which were not included in the plan (e.g. last minute tasks, breaks), and 5 hrs 46 min spent on activities that were included in the plan. In other words, 15.6% of the workday was spent working on tasks that were not planned.

Amount of work not completed by end of the day

Out of 7 hrs 44 min work tasks planned, tasks with an estimated duration of 1 hrs 43 min (SD = 116 min) were not started at all by participants. In addition, participants started working on average for 30 min on tasks estimated at 84 min that they did not manage to complete. Hence, they left at least an additional 54 min of work incomplete. The estimated duration of all work tasks left incomplete was 2 hrs 37 min of the 7 hrs 44 min planned, or 34%.

Time estimation bias according to the type of task

In the following section, the different groups of tasks according to the type of estimation bias are discussed, starting with correctly estimated, followed by underestimated and, finally, overestimated.

Correctly estimated task were scheduled meetings, seminars, reading research, and creating presentation tasks. Sixteen participants attended scheduled meetings. On average they planned to spend 78 min but actually spent 64 min in meetings. Examples of the kinds of meetings that participants attended are meetings booked by a student, meeting one's supervisor, meeting one's research group or team. Apart from scheduled meetings, six participants attended scheduled seminars which were planned as 130 min but actually lasted 110 min. Further, 11 participants planned to spend 68 min on reading tasks but later reported spending 66 min on this task. Example reading tasks included: reading an article prior to supervisory meeting, or reading a draft of students' work. Finally, three participants planned to spend 91 min but spent 77 min on creating presentations, such as preparing slides for their Viva examinations or for their lectures at university. Almost all of those tasks were related to an important commitment (e.g. lecture or a meeting) on the same or on the next day.

Tasks with underestimation bias were email and communication, and coding tasks. Twelve participants planned to spend 51 min but later reported spending 68 min on average on email and communication activities throughout the day. They reported hoping that they would have fewer emails than they actually received. Coding and data analysis was the other group of tasks that took longer than expected. Five participants planned to spend 90 min but actually spent 182 min on those tasks. The most frequent reasons for delays were either an unexpected code issue that needed urgent attention or because they forgot to pre-process data.

Tasks with overestimation bias were writing and planning research tasks. Twelve participants expected to work on at least one research writing task with an estimated mean duration of 119 min but later reported spending 78 min on this task. Some participants defined a clear objective in their writing tasks, such as *fix introduction comments*, whereas others aimed to spend a certain amount of time on their research papers or chapters, such as *write research report for 2 hours*. Next, five participants planned to do a planning or decision-making task with a mean duration of 63 min but actual spent 19 min on this. For instance, they wanted to plan their month ahead, plan how to approach a writing task, or decide on a research direction. Overestimated tasks tended to be left unfinished unless they were related to a very pressing deadline. Participants often explained that these tasks were left incomplete because they tended to be cognitively demanding tasks.

3.3.2 A typology of common reasons for delays

All participants had differences between planned and actual activities. They were invited to discuss and reflect on those during Interview 1. Many noted that they *would not have noticed where time went unless they kept the diary* (P7). Reflecting on the diary revealed insights about tasks they spent longer than they thought.

The thematic analysis of interview 1 resulted in 10 types of common planned work slow-downs grouped in four themes: *Preparatory work, Breaks, Requests and Fatigue*. Those themes describe the events that caused discrepancies between how people aimed to achieve the tasks that they planned, and how they actually executed them. Table 3.4 summarises this analysis and provides examples from the data. Most of the identified types of slow-downs were due to imprecise and vague planning, with a minority being due to external interruptions during execution of planned work tasks.

Some participants viewed the omissions they made in their plans as something they could have better accounted for. P10, for example, noted that *I do not normally estimate the duration of my daily tasks. I make a to-do list and do not check the accuracy. I should do it because I am probably too optimistic*. Similarly, P8 concluded that he considered starting to account for breaks in his plans: *Some people may need less breaks than me but I'm not one of those. So, this means that I need to consider my breaks when I make my morning plan*. While appreciating that they could have planned more accurately, some participants noted that *plans were open to change*, and they did not mind that. Participants' view on their willingness to improve the accuracy of their plans is given more consideration in the next section.

Preparatory work

Preparatory work was a theme about doing extra tasks which emerged during the day in order to achieve the one that was planned in the morning. Those events were the following codes: omitting a necessary step, deciding to gather more information or deciding to organize information.

Group	Type	Description and examples	N
Preparatory work	Omitting a step	People forgot to include in the plan a necessary step for successful completion of a task, such as preprocessing data, doing references in papers, commuting time.	9
	Adding more information	People realized that they need more information for completing a planned task. For example, they decided to read more literature before writing a review or postponed a task while gathering enough evidence to make a decision.	4
	Organising information	People talked about an unplanned activity to make sense of new information, such as structuring a piece of writing, organizing library or doing a mind map of the literature.	3
Breaks	Social	People described taking a break to socialize with their colleagues. They wanted to get updated about other's work, to give or receive advice or to improve their mood.	7
	Physical	Participants described a small break to take a step back from their work. During physical breaks people walked around, tidied up their desk or got a coffee while thinking about work.	4
	Non-work	People reported taking a non work related break. During this time, they did a personal task or checked their social media accounts.	4
Requests	Task	People responded to a task request. For instance, they did an unplanned task after checking their email or they did a favor to coworkers.	5
	Meeting	People accepted a last minute invitation for a meeting or their meeting was cancelled.	3
Fatigue	Multitasking	People spoke about having to deal with several mentally taxing tasks. As a result, when they moved on to the next task on their list, they felt more tired than expected. For example, they had to deal with several urgent tasks and could not complete planned research work later.	6
	Monotasking	People planned to work on one cognitively demanding or very repetitive task and reported being too tired to work as efficiently as they imagined they would do. For example, they planned to spend a certain amount of hours writing but felt fatigued and switched to doing emails.	2

Figure 3.4: A typology of common reasons for delays in planned daily work based on Interview 1. Table shows description, example and grouping of each of the ten types identified. N = number of participants

A very frequent reason for delays in the planned duration of tasks was omitting a necessary step required for its completion. For instance, some participants reported forgetting about the required preprocessing and cleaning of data: *I realized that the data was a bit messy and I had to preprocess it before analysis* (P4). Others reported forgetting about the task of doing references for papers. They said that they simply did not remember it when making the morning plan: *I was planning to spend 30 minutes on finishing the paper but then I remembered that I had to do corrections in the references I just forgot to put it in the plan* (P14). Another code was related to gathering more information. For instance, participants realized they needed to read more literature in order to be able to complete a planned task: *I was not planning to read this literature but it seemed relevant to my review and I spent some time on it* (P18). In other cases, they decided postpone a task while gathering more information: *I thought I would write several paragraphs in the proposal but then I was not sure where it was going and how to write them so I had to go back and read more papers*.

The third code in the theme of Preparatory work was related to organizing information. Participants decided that they had to organize the information they were reading or writing about before moving on. For instance, they decided to organize their literature libraries: *I decided to sort out my papers in Zotero and that is why the review took longer* (P11). Others decided to make a structure of the piece of writing: *I was not planning to but I made a mind map of the literature after reading the papers. Then, I moved on to editing my literature review* (P6).

Breaks

Breaks was the second theme that was identified in Interviews 2. When reflecting why planned durations of work tasks differed from diary reports, majority of participants spoke about taking breaks while working. Some of them reported breaks in their diaries while others did not. Those who did not report breaks often remembered that unreported breaks made the completion of a task longer than expected. There were three different types of breaks: social, physical and non work.

Social breaks were when participants reported socializing with their colleagues in the corridors or during tea breaks in the common room. Those breaks helped participants obtain friendly advice. Informal conversations during those breaks were often work related such as discussing career choices: *We had a chat in the corridor. We discussed the job interview [which I just had] and she gave me advice.* (P5) Sometimes social breaks occurred simply to receive or give social support to coworkers who were challenged with a difficult task or a tight deadline: *I had a difficult day because I was struggling with analyzing my data I took a tea break in the hub to socialize and chat.* (P12)

Physical breaks were when participants described taking small breaks to take a step back from their work. During physical breaks people walked around, tidied up their desk or got a coffee. Participants shred that they needed time to think about how to solve a work task or to make a decision about what to do next. For example, P2 described taking mini breaks to help with putting together a presentation. P2 shared: *So I took a few breaks to empty drawers and it took me 15 minutes.* (Interviewer): Was there any specific reason? P2: *I was making the slides and I had to think about some things, structure, what exactly I was going to do and I had to take a step back and think about what is that I need to do so I took mini breaks to think about what I should do as next steps.*

Other participants realized that physical breaks get omitted when planning. P20: *I did not realize that I take that many breaks during writing [referring to a total of one hour spent in breaks. I often need to refresh my mind and to think about the problem I am trying to solve. Some people may need less breaks than me but I'm not one of those. So, this means that I need to consider my breaks when I make my morning plan.* Non-work breaks were when a participants did personal tasks. For instance, they checked their social media accounts. P9: *I take a lot of breaks. I did a few personal small tasks for about 10 minutes each and I also fixed my Dropbox account.*

Requests

Work request was another theme in the data. Some requests were related to accepting new meetings on the same day and other were about doing a new task. Meeting request were sometimes related to meeting a student and were requested by people who participants felt obliged to say yes: *A student contacted me. She wanted to discuss her research and asked about how she could do a PhD with me the request came from somebody important to me* (P18). P11 needed to accept a meeting because her team wanted to satisfy an unhappy client. *The client meeting... it was supposed to be one hour they were unhappy, team miscommunication ...We then scheduled an internal meeting and it took time to figure out what to do to keep them happy.*

Tasks was another type of request participants received. They had to do a new task which was requested by others. For instance, they received a task request during a meeting with their research team. *He asked if anyone had notes from that meeting and I said that I could send him mine if I had* *time. I did this in the afternoon.* (P9) Others did administrative tasks following an email: *I am part of this group and we need to deal with the module so they sent me an email to do something related to the module on Moodle* (P20).

3.3.3 Fatigue

Fatigue is a theme related to leaving tasks incomplete due to having no energy to work any longer. They often did not finish all of their tasks not just because of lack of time but also because they were feeling tired. Participants had different preferences for the variety of their tasks within one single day. When this preference was not met, they reported feeling fatigued. Multitasking was one of the codes related to fatigue. Some had many projects to work on during one single day and switching between them proved to be more difficult than expected: *I was too tired at the end of the day to do research work. The day was draining with all those issues that came up* (P13). Monotasking was another code related to fatigue. On the contrary to having too much multitasking, in some cases participants planned to work on one long and attention demanding task and reported being too tired to finish the whole task. *I did not have enough stamina to spend five hours [as planned] on the paper so I started doing my emails* (P8).

3.3.4 Personal planning strategies

The thematic analysis of interview 2 resulted in four themes related to the strategies participants used for their planning: minimal, daily, weekly and multi-level (see Table 3.1). Each strategy is presented and illustrated in turn below.

The minimal planning strategy

Some participants did not use planning tools systematically and were not in the habit of estimating how long their tasks will take. They referred to their calendar, wrote on post-it notes or on a whiteboard to keep track of their deadlines. They enjoyed being able to flexibly react to tasks *as they come*.

Participants expressed that their minimal planning habits were working *well enough* for their current workload. The minimal planning strategy allowed participants to avoid spending time on constant planning and re-planning, and to avoid *feeling bad about [oneself] when things come up and mess up the plans* (P4). As P4 mentioned, *I can react flexibly to what's most important but if I make a plan* *and something else comes up and then I'm a bit annoyed [...] Planning feels a bit tedious.* In addition, minimal planning was used as a temporary strategy prior to deadlines because *it is very clear* what the next task is (P14).

Participants noted that minimal planning strategy had limitations. P15 shared that at the beginning of the PhD she was not making specific daily plans. She expressed that: *It's very unspecific and in general you waste more time.* P8 noted: *I took it easy in the last months and I've noticed that work explodes before the deadlines [...] you can also overload yourself with work [when you don't plan].*

Participants who had a minimal planning expressed that they would benefit from a system which measures their energy and recommends a suitable task to match their focus levels. They expressed a need to predict optimal time for tasks, eliminate distractions or know when to stop working in order to avoid losing energy on the next day.

The daily planning strategy

Some participants planned their tasks once a day in a to-do list. They did not use any other type of task list (weekly or termly). Majority of participants in this group used general tools, such as calendaring apps, word documents, note-taking software (e.g. OneNote) or sheets of paper for their daily plans. Only one participant used a dedicated to-do list software (Microsoft To-Do). In addition, participants who had different responsibilities related to research and teaching used a combination of daily planning tools. They did so because they wanted to avoid forgetting important tasks and to capture tasks from different sources. However, having to manually transfer a multitude of tasks from one placeholder to another could be inconvenient. P15 shared *I use a combination of things. It's slightly annoying because [the calendar] duplicates what the to-do list does.*

Daily planning allowed participants to have a more manageable and realistic list of tasks because they could identify all components of the tasks they had to do. Daily planning required some estimation skill. For instance, P2 noted *I make a to-do list at the beginning of the day, sometimes at the end of the day and I update it when I do a task. I also break down my tasks into smaller tasks while I'm working [...] It's easier to have daily lists because you can see what can be done in one day. One very long [monthly or weekly] list would be overwhelming.*

Planning strategy	**Benefits**	**Limitations**
Minimal planning	Saves time and effort to plan Allows picking the next task according to mood and energy Gives sense of flexibility	Time can be lost without noticing Deadlines can be stressful Can lead to unbalanced workload Tasks can be forgotten
Daily planning	Improves time estimation skills Supports work detachment Reduces worrying about tasks Gives a sense of achievement Helps identify daily priorities	Does not give a sense of direction Tasks duplicate across tools Requires consistency Difficult to create enjoyable plans Hard to plan far ahead
Weekly planning	Gives a sense of direction Allows assessment of weekly workload Reduces worry about deadlines Gives sense of flexibility	Not accurate or precise No sense of achievement Smaller tasks can be forgotten Difficult to multitask
Multi-level planning	All benefits of daily and weekly Allows workload assessment Enables planning time away from work (e.g., days off)	Time consuming and effortful Re-planning can be stressful Challenging to find suitable tools May encourage overworking Collaboration delays

Table 3.1: Summary of personal planning strategies together with their strengths and limitations based on the data from Interview 2.

Daily planning also helped to reduce worrying about tasks and was linked to feelings of satisfaction: *From my list I say these are the priorities for today and I cross them out once they have been done. I enjoy when I cross tasks [...] If I write things down I don't have to worry about remembering it. I get easily stressed about the things I have to do. I don't want to do things the minute before the deadline.* (P7).

Participants using this strategy often expressed that it was important for them to have a consistent schedule. They would not work according to their energy: *In the evening sometimes I'm quite productive but I don't allow myself to work[...] Tomorrow I will follow the list I made today after work. It feels quite nicely to know what comes next.*

Finally, daily planning had some limitations. P13 noted that her strategy was not working *anymore* since she got very busy in her PhD: *I tend to overcharge [my paper planner] every day which means that things go to the next day and the next day... And this is why I think it's ineffective.*

The weekly planning strategy

Some participants made weekly plans. This strategy was used as a way to confirm that progress was made toward long-term goals while urgent deadlines were under control. It also allowed flexibility about when tasks can be done during the day. Participants used pen and paper or a dedicated software (e.g. Trello) to make weekly plans. Weekly planning did not require accurate time estimation. Participants who used weekly planning tended to work on one project at a time, for instance, they had longer period of time to focus on preparing a lengthy report (P1). These participants expressed that it was challenging to switch back to their primary task, such as writing, after doing other unrelated shorter tasks, such as attending meetings. They preferred to block out days for single tasks.

To illustrate, P1 listed and reflected on his planned tasks on a weekly basis. He had just started to use a tool (Trello) which he updated at the start of his work week. He noticed that his planning was not accurate but he kept doing it because the activity was helpful for his motivation at work: *[Weekly] I go Trello and in my inbox and think about what I need to do, the things which are fixed and the things I can push as far as I can or I can delegate to someone else. [...] I then reflect on how that week maps out in terms of milestones, how it will help me get my PhD [...] I get a vague idea of the main things that need to happen but I remain flexible. [...] The plan works at the beginning of the week, at the middle of the week is less accurate and on Friday I am really behind. Planning is not accurate, I* *feel it is therapeutic.*

P12 expressed she wanted to remain flexible and work whenever she felt like working. She would plan that *this week and that week were for literature review* and at the same time she *did not want to be controlled by a plan* and she was happy to *work hard two days before the deadline*.

Some participants had stopped using weekly planning because they had realized that they could lose time without noticing. P15 shared: *Early on in my PhD I would decide to spend one day on writing and I wasted a lot of time [...] Instead you should have more chunks of work and make smaller tasks with deadlines.*

Weekly planning was not effective for those who suddenly started managing complex and varied workload. P20 was a newly appointed lecturer (assistant professor). She shared that she could not use her typical planning strategy and she needed to find a better one: *Normally I had been in a habit of once a week paper to-do lists and then I became busy and had no time to make a plan. Last plan I made was [two months ago] and with tasks that cannot be achieved in a week like write a chapter for my book. I also have 20 min tasks and since the [date two months ago] I have not been able to use it at all. I tried to hold things in my head and respond through my email as a to-do list and I've been really overwhelmed in the last month.*

The multi-level planning strategy

Some participants made task lists at different levels of granularity. For instance, they made a master project list each month or week and daily lists alongside. They did this by adapting generic tools or by personalising a dedicated tool, such as Trello. The use of task lists at different levels of granularity was driven by the need to synchronise multiple schedules and projects deadlines. They wanted to be able to anticipate their tasks and to *advance their understanding about how to balance different responsibilities and associated deadlines* (P8). Participants who used multi-level planning tended to have a complex workload and a variety of deadlines and projects.

P16 shared that keeping a daily list alongside her other lists helped her estimate more accurately the duration of her tasks in the long-term future: *I have all levels: daily, weekly, monthly. I can go back and check how much time it took me to do something like a report when I am planning to do it [in the future]*. A similar benefit was the multi-level planning allowed participants to complete tasks more efficiently. P8 shared: *[Planning] is a lot of work but I am much more organised and able to get* *things done more quickly.*

The gained efficiency from multi-level planning had a time cost. One of the challenges with multi-level planning was that it took significant amount of time to do. Participants felt that there were no tools to help them automate the process of figuring out how much time they had spent on different tasks. P8 noted that: *One full day every month goes to figuring out deadlines and calculating the time for all my different commitments and another full day on calculating what I did in the past month.[...] I wish there was a tracking tool to do this for me.*

Another challenge with having multi-level planning was that participants expressed that not completing day's work tasks would negatively affect subsequent days. As P11 expressed *I set the main tasks [...] depending on how many I can fit, it's usually 4-5 things, and depending on how many I have ticked off on the previous day. Yesterday I did not tick off everything so today I have double the amount of tasks and this is something that I need to improve.* A related challenge participants expressed was that even though they were more organised, they were also more likely to *stay later [in the office] than usual to meet the plan* that they had set for themselves. Finally, some participants expressed that even if they managed to plan out their time effectively, they were often waiting for others to reply and do their part of the work on collaborative projects: *Usually it's others that slow down my work* (P16).

3.4 Discussion of Study 1

Study 1 set out to examine why academics so often struggle to plan tasks accurately. The findings showed that inaccuracies were not random but patterned: certain task types were more prone to underestimation, and delays were often explained by contextual factors such as vague task definitions, hidden subtasks, and frequent interruptions. The qualitative data further revealed how participants interpreted these mismatches and described the strategies they used to cope. In this discussion, I situate these results within psychological accounts of the planning fallacy and HCI research on personal task management, before outlining their implications for the design of digital planning tools.

3.4.1 Time constraints and optimism

The first research question in this study examined the ways in which work planning is inaccurate. Suchman [80] claimed that plans are imperfect because in the large they cannot encompass all possible outcomes and contingencies. In line with this view, we found that daily task planning in knowledge work is often inaccurate. Participants expected to work one hour longer than they actually did, and 34% of planned work was not completed by the end of the day. To compare with previous literature, [3] found that 27% of tasks were not completed. These findings are in line with general optimistic attitude in planning [81].

The findings make a substantial contribution going beyond previous literature by showing that some types of information work tasks are more prone to estimation errors than others. On the one hand, the duration of time constrained tasks, such as tasks related to scheduled events and those with clear deadlines, tended to be estimated correctly. On the other, the duration of less time constrained tasks, such as tasks that were not scheduled and those without clear deadlines, tended to be estimated incorrectly. There could be several possible explanations for this finding. First, Redaelli and Carassa [82] argue that people follow plans in so far as they are *instructions* for situated actions; people are more likely to follow clear and instruction-like plans. These instructions hold them accountable by clearly indicating what is correct and what is wrong to do. From this perspective, time constrained tasks (e.g. commitments related to events) may have significant impact and hence increase a person's sense of accountability. Second, time-constrained tasks were also clearly defined and specific as opposed to being open-ended tasks. People can easily change plans about less time constrained tasks later on during the day depending on whether other more time constrained tasks suffer delays or get canceled.

Previous research has examined time estimation biases by contrasting broad groups of tasks (e.g. information-related tasks compared to people-related tasks [15]. The current study goes into greater detail than previous research in this space, by showing that there is a divergence in the direction of the time estimation bias between different kinds of information-related tasks. In particular, writing, planning and decision-making tasks were found to be more error prone *in terms of overestimation*, while coding and email tasks were found to be more error prone *in terms of underestimation*. This divergence in the direction of the time estimation bias can be explained in terms of optimism about task efficiency. Doing a great job sometimes means speeding through a task (e.g. how quickly an email will be read and replied to) and sometimes quality work is about spending more time on a task (e.g. how many iterations of the paper there is time for). It is challenging to perform writing and planning in less time than the time they inherently require because compressing them can lead to a lower quality result. Existing literature describes attempts to support workers to spend *more* time in deep work, e.g. [83]. In comparison, coding tasks and emails could be done *quicker*, and probably in rare occasions are done quicker than expected, without losing quality of results; for examples of previous research that have explored ways to speed up performance in coding and email work, see: [84, 85, 86]. Therefore, the divergent direction of time estimation bias can be explained in terms of optimism about achieving better performance on different types of tasks.

The observed divergence in the direction of the time estimation bias between different tasks extends the literature on the temporal elasticity of work [87, 15]. Previous studies have put forward the idea that workers tend to compensate for the slower execution of information work by speeding through face-to-face meetings. Evidence suggests that knowledge workers today spend less time in face-to-face meetings and more time on information work, such as email [88]. The temporal equilibrium identified in previous research may look different today. While workers three decades ago tended to spend more time on information work than expected and less time in meetings than planned as a way to achieve balance [15], we found that workers today may lack opportunities to achieve balance. The findings suggest a lack of available time for writing, planning and decision-making tasks due to significant delays in all other types of information work. Other types of information work, such as email and coding tasks, unlike face-to-face meetings, is probably more challenging to speed up even though workers optimistically plan to do so in the morning. This observation suggests that in the long-run writing, planning and decision-making tasks are less likely to be done to the standard they were planned at due to the direction of optimism bias in information work tasks.

3.4.2 Vague planning and slow-downs

The second research question in this study aimed to examine the reasons for discrepancies between plans and actions. The typology of reasons for delays in planned work indicates that most participants made plans that lacked detail. Specifically, participants did not factor in enough time for preparatory work tasks, breaks, requests from others and lost time due to fatigue. While previous literature has classified the interruptions people experience during tasks [89], there were no previous studies to classify the events that prevent people from achieving their daily plans.

The analysis of delays suggests that there were estimation errors not only because participants overestimated the hours they would spend working, but also because of the optimistic and vague way they planned their tasks. Participants envisioned the end goals associated with the tasks they executed, but often did not give sufficient attention to all of the steps implicated in achieving that goal. For instance, when they aimed to analyse data from sensors, they did not think about eliminating noise from the data in the plan. They also did not plan time for other time consuming activities that were not related to their goals, such as breaks and responding to requests of others. Hence, day-to-day work planning lacked detail; it focused on what participants wanted to achieve rather than what was likely to happen. Interventions which ask people to plan in detail all of the steps included in a task are found to reduce the planning fallacy [69]. For example, previous research has recommended that digital calendars implement a function to allow users to allocate a flexible preparation time before their events, such as travel time [9, 15]. The current study contributes to the literature about the optimistic plannin bias (planning fallacy) [90, 91, 92] by suggesting that there are at least two mechanisms behind it: participants were optimistic about the time they would spend working and were optimistic about how efficient they would be at completing work tasks during the limited time they had available.

It has to be noted that receiving unplanned requests was relatively uncommon. Most of the types of delays could be categorised as planning failures rather than as external unpredictable work. As the diaries found, only 15.6% of activities at work were not planned, part of which were breaks. Majority of previous studies on planning at work has concluded that planning is unreliable due to unexpected external events. In hospitals, for example, 50% of executed work is not planned and due to external events [82]. This study finds, in contrast, that planning accuracy is within people's control at least in some knowledge work environments. Academics and researchers in academia are known as independent workers who are in charge of their long-term projects and goals. In those contexts, planning can become more reliable if workers acquire better planning skills, e.g. by making more detailed plans. Reliable planning is achieved through experience and regularity [7, 93]. Therefore, we put forward that accurate planning in some knowledge work contexts is achievable.

3.4.3 Accuracy of planning strategies

The final research question aimed to investigate whether some planning strategies were more effective than others. The findings showed that it was important for participants who engaged in detailed planning of their work activities (e.g. multi-level or daily planning strategies), to have an accurate idea of how long tasks would take to complete. In contrast, those who adopted a minimal or weekly planning strategy, were largely unaffected by their time estimation failures. This would suggest that poor accuracy in time estimation in planning was *intentional* for some of our participants. It could be argued that a minimal or weekly planning strategy does not support accurate time estimations. Instead, these strategies allow workers to be spontaneous and flexible when deciding which task to execute next. Previous research by Haraty et al. [55] found that Make-doers have similar planning habits to what is described here as minimal planning. Make-doers report the busiest of all three groups in the study. While similar evidence in terms of business was not found, the current study can offer an explanation for this finding in Haraty's work [55]. One of the limitations of minimal planning is that it is linked to pressure before deadlines. Minimal planning is prone to estimation biases. It may cause delays in delivering work projects and time pressure prior to deadlines. The findings also showed that some workers use minimal planning temporarily prior to deadlines because the tasks that they focus on are very clear and do not require any planning. This can explain why on average this strategy is linked to feelings of business in Haraty's work [55].

This study also suggests that people's planning strategy is sensitive to their workload. Interviews with participants showed that when those with minimal or weekly planning strategies started to manage a higher number of projects and became in charge of other people, they would switch to a daily or multi-level strategy instead. These strategies emphasised time estimation accuracy. In addition, other participants reported switching from multi-level to minimal strategies. This finding leads to the conclusion that workers change their strategies over time. Planning is, therefore, strategic and adaptive behaviour. It is flexible and responsive to job demands. Planning strategies may hence be influenced by contextual factors to a greater extent than they are influenced by stable individual differences such as orientation towards planning in general [94] and preference towards personalisation in tool use [54]. More research is needed to explore the extent to which individual differences in the use of planning tools and strategies are context-dependent and the extent to which they are influenced by stable factors over time such as personality.

Finally, it has to be noted that participants reported difficulties in finding the right tools for multi-level planning compared to the other planning strategies. Participants who used a multi-level strategy reported using other strategies in the past, such as weekly or daily planning. Multi-level planning required much more time than other planning strategies, and it also required participants to search for appropriate tools to track and plan tasks on different levels of granularity, such as yearly, monthly and daily. This finding points to a gap in the tools that support multi-level planning for knowledge workers. This gap may be due to a lack of tools that provide ways to plan effectively on different levels of granularity. For instance, there could be a mismatch between users models of how daily and longer term plans are coherently integrated and the functionalities that planning tools offer to differentiate between daily and other forms of planning. Future research should explore how existing technology can be improved to meet the needs of workers with multi-level planning.

3.4.4 Limitations specific to Study 1

The study has several limitations, which we critically consider here. First, asking participants to make a plan in the morning could have an influence on how they spend their time later on. Even though the morning plan was given to the researcher in the morning and participants were instructed to behave as usual during the day, they might have had a recollection or a copy of their plans. Hence, they might have put more effort than usual to follow their plans. Nonetheless, the current study still found time estimation errors. It is likely that those errors would be higher when people are not explicitly asked to make a morning plan.

Second, we measured the amount of time participants expected to spend on a task as opposed to the amount of time a task would take to complete. The latter can span over several days. Investigating delays in task completion can further our understanding about estimation accuracy in planning. This can be done with a longitudinal diary study. While the one-day augmented diary followed previous work [15] it could still be argued that a longer study would yield better insights. To address this limitation, we asked participants to chose the day of the diary specifically asking for a typical day with a balanced amount of different tasks.

Third, this was a mixed-methods study with 20 participants. It is possible that the typology we present is not complete. Subsequent studies can extend the findings with a larger sample and a different design motivated by the findings in this study. Moreover, the sample of participants included PhD students, postdocs and lecturers. The insights therefore are industry specific and can be applied to other knowledge workload with similar types of workload. Future research can extend this work with other groups of knowledge workers.

Finally, the current study explored in depth the subjective experience of inaccurate planning. One of our aims was to investigate whether inaccuracies at planning are perceived as troublesome. While this approach yields valuable findings regarding the lived experience of planning failure, it does not provide detailed insights about all possible factors that may underlie planning inaccuracies. Some of those factors include the various technology participants use for planning, different work domains within each job sector, the time of the day (morning or evening) for which a task is planned, and other task characteristics such as priority or importance. Future studies may explore those factors with alternative study design, for example, an experience samping approach over the course of several weeks, as in [3].

3.5 Summary of Study 1 and bridge to Study 2

Study 1 examined how, when, and why task planning is inaccurate in academic knowledge work. The findings showed systematic optimism: participants planned to work around one hour more than they actually did, and 34% of planned tasks were left incomplete at the end of the day. Task-level analysis revealed distinct patterns of error across activity types: cognitively demanding tasks such as writing, planning, and decision-making were more often overestimated, while routine activities such as email and coding were typically underestimated. These discrepancies suggest that delays accumulate across tasks in ways that displace complex work, undermining long-term progress and quality.

Most inaccuracies were not the result of external interruptions but arose from vague, aspirational, or incomplete planning. This indicates that accuracy is at least partly under individual control and could be improved through support for task structuring and more realistic time estimation. Participants also described distinct planning approaches, ranging from minimal to multi-level planning, adjusting their style depending on workload and role demands. This underscores the need for flexible, personalised forms of planning support.

At the same time, planning practices are sensitive to wider context. Major disruptions or transitions, such as shifts in responsibilities, workload, or working conditions, can prompt individuals to abandon, adapt, or rebuild their routines. Understanding planning accuracy therefore requires attention not only to the strategies people use, but also to the conditions under which they change them. Study 2 addresses this by tracing how academic knowledge workers adapted their planning during the early stages of the COVID-19 pandemic. It shows how participants responded to disruption by disengaging from old routines, experimenting with new approaches, and reflecting on their effectiveness, offering insights into how technologies might best support people at moments of transition.

3.6 Introduction to Study 2

This study addresses the first research question of the thesis: *What are the main challenges in accurate task planning at work?* While Study 1 offered a snapshot of how day-level planning inaccuracies manifest in everyday academic settings, it captured planning behaviour at a single point in time. It revealed that many planning failures stem from vague and optimistic expectations, but also that inaccuracies are often within users' control, especially for knowledge workers with high autonomy like academics. However, the study did not examine how individuals adjust their planning practices when their working context shifts dramatically.

To address this gap, Study 2 complements the findings from Study 1 by investigating how planning behaviours evolve over time, especially during moments of disruption. Transitions such as job changes, increased workloads, or shifts in life circumstances are known to prompt re-evaluation of planning strategies and tools [54, 2]. These periods present both challenges and opportunities: people may disengage from planning routines altogether, abandon their usual tools, or develop new adaptive strategies. Designing planning technologies that are sensitive to such transitional contexts requires a deeper understanding of how and why planning behaviours change in the face of disruption.

The COVID-19 pandemic, which began in early 2020, created an unprecedented moment of upheaval in knowledge work. The sudden shift to working from home (WFH) disrupted the structure of daily routines, blurred the boundaries between work and non-work time, and introduced new logistical and emotional challenges. This disruption provided a unique opportunity to examine how academic knowledge workers respond when their existing planning habits break down. In particular, the study investigates whether planning disengagement occurs, how workers attempt to restore control, and what kinds of strategies and supports are most useful during such transitions.

This study explores how individuals adjusted to WFH during the first COVID-19 lockdown in the UK, and what impact this had on their planning routines. It asks two specific research questions to address the main research question (see above): First, what kind of support do people need to advance their own planning routines? Second, are there any planning strategies and tools that workers find helpful during a disruption? To answer these questions, a longitudinal qualitative study was conducted with 15 participants working in academia in the UK and US. Over a period of three months (AprilJuly 2020), 76 remote interviews were conducted, offering a detailed view of how participants' planning behaviours shifted as the lockdown progressed.

The findings reveal widespread planning disengagement in the early stages of lockdown, with many participants reporting a loss of motivation, breakdown in routines, and uncertainty about how to plan under new constraints. However, the study also documents how some participants, gradually developed new planning strategies to regain a sense of control and productivity. Notably, the findings show that planning in detail, even during uncertainty, can foster feelings of agency and improve work engagement. Moreover, participants who particularly reflected in depth during the interviews on their use of time were more likely to adopt new strategies that supported both planning and productivity.

This study makes three key contributions. First, it provides a rich account of how the pandemic disrupted existing planning routines, demonstrating how planning disengagement emerged and evolved over time. Second, it identifies two concrete planning strategies: breaking down tasks and manual time tracking, that helped participants regain a sense of control by improving the realism and accuracy of their plans. These strategies reflect organic efforts to mitigate optimistic planning bias under challenging conditions. Third, the study offers design insights suggesting that planning technologies should not only facilitate the mechanics of planning, but also help users re-engage with planning behaviours during periods of disruption or low motivation. By showing how individuals independently arrived at effective, bias-reducing strategies, this study highlights the potential for technology to intentionally scaffold such approaches. These findings lay the conceptual and empirical groundwork for the next studies, which systematically investigate how psychological debiasing strategies can be translated into digital interventions, and how well existing task management tools support these strategies in practice.

3.7 Method of Study 2

3.7.1 Participants

Fifteen participants took part in the study (see Table 3.2). They were academics and early career researchers at UK (N = 14) and US (N = 1) universities (3 x lecturers, 2 x post docs, and 10 x PhD students). They were all working from home during data collection as a result of the COVID-19 pandemic. Participation was voluntary. The study was approved by the university ethics committee.

3.7.2 Design and procedure

Lockdown in the UK started officially on 23 March 2020 and on different dates in March in the US depending on the state. However, all participants were working remotely by 16 March. Data collection began on 6 April. Semi-structured interviews were conducted each week for an average period of five weeks with each participant. Eight participants joined during the second week of April. The remaining seven participants joined gradually in the following weeks. All participants had joined the study by the beginning of June. Data collection finished in mid-July.

Each interview followed a simple semi-structured script designed to explore participants planning practices during lockdown. Interviews typically began with broad questions about how participants had organised their work in the past week and whether anything had changed since the previous interview. Participants were asked to describe the tasks they had been working on, how they had planned them, and any difficulties or successes they had experienced in following their plans. The researcher also asked whether participants had adopted any new strategies, tools, or routines, and whether they had discontinued any previous approaches. When relevant, the interviewer probed for concrete examples, such as how participants estimated task durations, how interruptions affected their plans, or how they adapted when tasks took longer than expected. Each week, follow-up questions were developed based on the participants previous interview in order to track changes and deepen earlier insights.

Interviews were conducted online through Microsoft Teams and were audio recorded. Recordings were transcribed with Trint (`https://trint.com`) and edited manually by the researcher to correct for mistakes. Data were then thematically analysed with Nvivo 12.

3.8 Results of Study 2

The data consists of 76 interviews with a total duration of 16 hours 30 minutes. Each participant participated in five weekly interviews on average (range from 4 to 6), with a mean duration of 13 minutes.

#Ppt	Gender	Occupation	Household	Routine	Primary Tool	D	New Strategy
P1	F	PhD student	Family (no children)	Weekly	Paper diary	Y	BDT
P2	M	Postdoc	Alone	Weekly	Paper diary	Y	None
P3	M	PhD student	With parents	Weekly	Word	N	BDT
P4	M	PhD student	Alone	Minimal	Notes (Mac)	Y	None
P5	F	PhD student	With partner	Daily	Notebook	Y	BDT
P6	M	PhD student	Family (children)	Minimal	Paper	Y	None
P7	F	PhD student	With partner	D+W	Word	Y	MTT
P8	M	PhD student	Alone	Minimal	Paper	N	None
P9	F	PhD student	Family (no children)	Weekly	GoodNotes5	N	BDT
P10	F	PhD student	Family (children)	Minimal	Paper	Y	None
P11	M	Lecturer	Family (no children)	Daily	Trello	Y	BDT
P12	F	Lecturer	Family (children)	Minimal	Paper	Y	None
P13	F	Postdoc	With partner	D+W	Trello	N	MTT
P14	F	PhD student	Alone	Daily	Paper diary	N	MTT
P15	F	Lecturer	Family (children)	Minimal	Word	Y	None

Table 3.2: Summary of participants backgrounds, planning routines and tools, reported disengagement from planning, and new strategies during the study. Pp = participant number; D = Disengagement from planning; D+W = Daily and weekly; BDT = breaking down tasks; MTT = manual time tracking.

3.8.1 Challenges brought by the lockdown

Poor ergonomics

Participants were struggling with physical pain and focus issues due to poor ergonomics:*The distance between my desk and bed is about a meter [...] Initially, I got to do 20 minutes of work and then I'd be needing to stand up and move around and stretch* (P2). Having a suitable work station at home had immediate positive effects: *Having a desk, monitor, mouse and keyboard has been really helping me a lot with productivity* (P9). When participants were experiencing a drop in work focus and motivation, they often re-arranged their home office. As P8 shared: *From the third weekend, I felt it was difficult to focus. [...] The first thing I did [to improve my focus] is I moved my desk to another place. And I re-arranged the room layout in my bedroom.*

Distractions by pandemic-related content

During the early stage of the lockdown, participants prioritized staying in touch with family and coworkers, and followed the news. Work focus was diminished: *It was just a burnout with the emails and all the messages, and I could not get anything done because it was just so much information at the same time. I was checking the news quite a lot* (P7). However as weeks went by participants noticed that their focus and concentration was naturally improving. P4 noted that it *was easier to go into a work mode in the mornings*.

Blurred temporal and spatial boundaries

Participants acknowledged a need to better separate their work and non-work lives. *I think I need a clear separation between work and rest [...] not only about the spatial work space but also about the temporary dimension* (P8).

Commuting was no longer serving as a natural separation between work and non-work times. *When you're commuting you start getting in the mindset of working and then when you come back, you start getting the mindset of relaxing* (P4).

There were no social cues to serve as reminder to stop working. *But because working life is so blended at the moment, sometimes it's difficult to stay focused when you don't have other people around you working. And if I start working, I might be working until 19:30 because other people aren't around leaving the office* (P2).

Family duties were highly unpredictable. *If a child starts crying, everything else has to be moved to accommodate that* (P15).

Participants shared that using the same devices for work and non-work activities was challenging. *I have always connected using my laptop with nighttime, like watching Netflix* (P10).

Feeling overworked and tired

Many participants had to change the directions of their research projects due to the new constraints of the lockdown which created extra work. *We've had a lot of issues around dissertations and people not being able to collect data and having to change the research questions and the data they would collect* (P15).

Participants who were living with their children experienced excessive tiredness. *and I do a little bit of work when the kids go to bed, but I'm just sick, exhausted that I want to go to sleep* (P6).

The added challenges of lockdown fatigue, social isolation, not having rest in the weekend, all made it seem like a regular day of work during lockdown was equal to a much longer workday before the lockdown. *There is nothing to look forward to in the weekend [...] I'm tired all the time* (P2).

Changing expectations for productivity

Many participants thought that the first lockdown would last for several weeks only: *We all thought this would last about three to four weeks and we'd go back to normal* (P10). Participants continued to work at a similar pace at home as before. *I've been working really hard during the week [...] I do the things that I used to do before everything happened* (P2). P8 shared that he followed the plans he made before the pandemic: *I made this plan before the isolation and in this first two weeks I was working quite efficiently*.

During the lockdown, many participants expressed low mood and a drop in motivation due to the realization that the lockdown measures would not be completely relaxed for months ahead. *I think it's starting to get to me, the fact that we are going to stay in this situation for a longer period* (P2).

Participants noted that they had put high expectations on themselves at the start of the lockdown. *...for the past six weeks, I've been working every day for at least eight hours. I'm just putting way too much of an expectation of myself* (P9). Many tried to relax the self-imposed schedules they had created. *I was very strict with myself waking up at 7AM and starting working by 8AM. I try to tell myself that it is fine to feel like starting work slightly later* (P13).

Participants' colleagues were understanding of the challenging situation and did not expect others to maintain the same level of productivity as before. *There is an informal understanding around the fact that people are not following their regular schedule. Say a student tells me "look, I've not been able to work on this", normally I might have pushed them a little bit but right now I don't push them the same way. I think all of us are working to that changed understanding* (P15).

At the end of the study, participants often recognised that lowering expectations for productivity was a key approach to the pandemic. *We need to cut each other a bit of slack. Being sensitive to each other's needs and what the environments will allow for and we'll get through it. We'll just have to adjust and adapt to new ways of working* (P12).

3.8.2 Planning routines

Relying less on planning routines

Table 3.2 shows participants planning routines and tools used during the study, and whether they reported disengaging from planning. Minimal routines refers to using planning only when necessary, daily routines refers to making daily task lists, weekly routines refers to weekly task lists, and daily plus weekly refers to making lists once a week together with daily ones. This classification follows the findings from Study 1 [77].

Throughout the study, participants frequently reported relying less on a planning routine. One of the reasons to not engage in planning was because they were focused on getting used to the new situation, such as adjusting ergonomically to the home office. They were distracted and had difficulties concentrating on work. Their planning was falling behind as they focused on small and urgent tasks: *It was quite hard to actually focus on my tasks. And if something was coming up, an email or someone texting you and getting in touch, that had the priority. I was not making a plan* (P7).

Participants relied less on their planning routines during extremely busy periods. Over the course of the study, many participants experienced high workloads due to extra teaching duties, or due to homeschooling. They minimized to scheduling activities and a very short to-do list. After a busy or heavily disrupted period passed, participants re-engaged with planning strategies oriented towards the long-term future.

P6: There is no time to plan, it's just doing. I squeeze in bits of work when I find a moment [...] Yesterday I woke up at 5am to try to get some work done. [four weeks later] *This morning has been my first day of planning. I'm making a new timeline for my research up until the end of my PhD.*

Lockdown schedules were inconsistent as participants were settling down into new routines. Often participants could not predict what they would want to do later on during the day which made planning more difficult. *I did not plan because I did not know how I would feel. I could not predict what I would want to do* (P11).

Participants said there was less need to plan when they were doing a repetitive task or working on a single project. *For what I've been doing this week, it's been pretty clear. I did not need my planner. Because I have this deadline of getting it done by Thursday* (P2). Having a mental plan was enough for them to remember what to do next. *On days I don't update the list it is usually because I only have to work on one task (rather than two or more), which I can remember, so don't need to offload it from* *my memory* (P5).

Relying more on planning routines

One participant reported relying more on planning routine because she found planning easier than before. She reported being able to make more accurate plans at home, given that there were no unexpected delays or interruptions. *I'm using my planning system more than before because it's more accurate. I don't have to take into account unplanned catch-ups or travelling from one place to another. Now everything is planned for and there is less error room* (P13).

3.8.3 New planning strategies during the lockdown

Even though some participants disengaged from planning during some periods of the the lockdown, they also started to engage in two planning strategies which they found helpful for their productivity: breaking down tasks and manual time tracking. Sometimes, they engaged in those strategies after first disengaging from their planning routines, and other times they did not disengage at all and used new strategies together with their existing planning routines. See table 3.2 for details about the strategies used by individual participants.

It is also worth noting that there were no themes in the analysis that referred to starting to use new planning tools, which reflects the observation that planning tools did not have high uptake rates. Participants used their existing tools for their new planning strategies.

Breaking down tasks

A planning strategy that several participants adopted during the lockdown was to break down their tasks into smaller units (than what they were already doing) while planning. In this way, they could set more achievable plans, feel better about work accomplishments, and increase their productivity.

P7 was already planning specific daily tasks but decided to make more detailed plans. This strategy meant she could learn how much time was being given to different tasks. *One thing I changed recently is that I was setting even smaller tasks for each day. That really helped, it can teach you how much to allocate for tasks in the future and how much work you can get done in a day* (P7).

P1 also decided to record all the small tasks in her weekly diary; previously, she did not consider small tasks to be *real work* but rather a distraction from getting real work done. This strategy boosted her sense of productivity as she noticed that the time she spent on the most important work tasks increased. *I think I put myself like very big goals and only the small ones get done. And the big ones, I never get them.* [two weeks later] *Instead of planning just to write, I tried to visualize everything else I had to do. This week my planner is full of small to-do things that were as simple as emailing people or planning this meeting. Those were things that I was still doing but I was considering them things in the way of what I had to do. Now they could give me motivation and I even wrote quite a bit even though I was having more things to do!*

Another way of breaking down tasks into smaller ones was by creating mind maps. P9 used her diary to map all the tasks implicated in the completion of her web application. This had immediate beneficial effects on clarifying the next actions she had to do. She also felt less overwhelmed. *I've created this tree map with subtasks of what I need to do exactly on the website because you can get overwhelmed. There are so many pages and functions. It was very difficult for me to identify which parts I need to do first. I was stuck for a couple of weeks.*

P11 started breaking his research work into project lists with specific tasks. *That's another novelty. I have a Trello board where I keep all my research projects, from master thesis for students to actual research projects with international collaborators. I look at each individual card to understand whether there is a next action. It's helping me to have a clearer idea of my progress. It helped move forward several projects which were stuck during the pandemic.*

Manual time tracking

Manual time tracking was a strategy in which participants kept a time log of their activities, often along with estimates of the time they thought a task would take to complete. Time tracking also helped participants become more accurate at planning. Sometimes they needed to be accurate because they had hard deadlines related to their projects. *I'm trying a new thing today. I estimate the hours a task would take me as I am realizing in the past days if I had like three things on my list, I got one and a half done. I started it because I have a deadline [in two months] to get my stuff in order. I need to get better at saying, OK, I'll work on these three things today* (P5).

Manual time tracking was used by participants who were feeling stressed and anxious about achieving their tasks. By learning how to set less ambitious goals, they could increase the chance of achieving their plans. *I wrote everything that I had to do and I have put it in front of me on my window.*

I have estimated approximately how many hours it should take because I underestimate that. I end up feeling very stressed because I am always behind (P14).

Some participants were time tracking prior to the lockdown and continued to do so during the lockdown. *I keep tracking my tasks because you can go back and see how much actual time things took. When I plan to do something in the future, I might think it will take two hours but I can go back and see that it actually took me three days* (P4).

3.8.4 Reflection as a driver of change

To develop new strategies, participants reflected on how productive they felt at work. Those who started using a new planning strategy had a less disrupted schedule, more stable environment, and, hence, more time to engage in reflection.

Participating in the study provided participants with opportunities to stop and think about the strategies they were using. *I think talking every week was a big motivation for me just because it forced me to think about [planning]. Sometimes weeks go by and then after three weeks I realize I have done nothing. This has been a good reminder of What have I done? What active steps did it take to make my situation better instead of worse? This itself was a good strategy* (P1).

The interviews reminded people of strategies they already knew would be useful: *It also led from the conversations we were having. I realized that we were talking about how creating subtasks was a good thing. And I knew that from before, but I wasn't really engaging in that. Ended up being pretty productive. I was very happy because the weeks before that I kept not meeting my objectives.* (P7).

Some participants had more opportunities for reflection than others. Those who had additional household duties such as homeschooling or housework noted that they lacked opportunities to stop and reflect. P15 shared that: *...with everything being blended in a way, you don't have those moments where you're forced to take some time out, when you're traveling or you're walking somewhere. Right now I feel like that reflective self time I don't have.*

3.9 Discussion of Study 2

Study 2 revealed how academics adapted their planning practices during the disruption of COVID-19. Whereas Study 1 showed that planning inaccuracies are patterned by task type and strategy, this study highlighted the flexibility and resourcefulness of planning routines when everyday structures were destabilised. Participants adopted new strategies (such as breaking down tasks and manually tracking time) that helped re-establish a sense of accuracy and control. In this discussion, I situate these findings within psychological accounts of reflection and HCI research on planning tools, and consider their implications for the design of technologies that can support planning under conditions of uncertainty.

3.9.1 The nature and drivers of planning disengagement

Consistent with prior research on remote work during the pandemic [95, 96], participants experienced significant disruption to their daily routines, work-life boundaries, and sense of productivity. Many struggled with poor ergonomics, inconsistent working hours, interruptions in the home environment, and the psychological toll of uncertainty. These factors negatively affected both wellbeing and work output and have been linked in prior studies to increased stress and multitasking strain [12, 96].

A key finding that builds on this backdrop is the widespread disengagement from planning activities. While task planning is generally seen as instrumental in managing time, deadlines, and competing priorities [4], many participants reported a loss of motivation to plan. Some stopped using tools they had relied on prior to the lockdown, while others stopped making structured plans altogether. This disengagement marks an important contrast to the findings from Study 1, where participants reported consistent engagement in personal planning strategies, even if those plans were sometimes vague or overly optimistic. Here, it was not merely the content or style of plans that was problematic, but the collapse of the very habit of planning itself.

This phenomenon can be interpreted through two complementary lenses. The first draws on decision-making theory. As Cox et al. [97] argue, people under time pressure often prioritize urgent tasks over important ones, relying on fast, instinctive decision-making processes (System 1) rather than slower, deliberative thinking (System 2). In the context of disrupted work, the urgency of reacting to incoming messages, adjusting to changing expectations, and managing family obligations may leave little cognitive bandwidth for the slower, more reflective task of planning ahead. Participants who did not engage in reflection reported falling back on reactive behaviours. In contrast, those who did reflect during the interviews were more likely to revisit and adapt their planning strategies. This suggests that reflection acts as a cognitive lever, helping people shift from short-term coping to more proactive and strategic planning.

The second explanation for disengagement relates to planning under uncertainty. As Study 1 demonstrated, time estimation is a core challenge in planning. During the pandemic, unpredictability in workload, task timelines, and personal schedules made accurate forecasting even more difficult. Some participants reasoned that planning was futile in such conditions, expressing a preference to see how the day unfolds. This aligns with prior research showing that when people cannot predict outcomes, they often disengage from forward planning altogether [81].

Importantly, the study also revealed that disengagement from planning may not be unique to the pandemic. Participants described past instances (e.g. starting new jobs, becoming parents, or returning from leave) when their routines were similarly disrupted and planning behaviours deteriorated. These insights suggest that disengagement is a recurring response to periods of upheaval. If we want planning tools to truly support people over the long term, they must be designed to address not just the mechanics of task tracking or scheduling, but also the motivational and contextual barriers that lead to disengagement.

3.9.2 The role of reflection in rebuilding planning routines

Despite the widespread disengagement, some participants regained their planning momentum over time. A major catalyst for this shift was regular reflection. Many participants used the weekly interviews not just to report behaviours but also to evaluate what was working, what had broken down, and what might help. In doing so, they naturally engaged in the kind of metacognitive activity that research on behaviour change highlights as essential for habit formation [98, 99].

The study found that reflection helped participants identify what they were missing without a plan, such as the ability to track progress or anticipate deadlines, and motivated them to try new strategies. These observations align with a broader body of work suggesting that structured reflection can promote behaviour change, particularly when embedded in routine practice or facilitated through digital tools. Recent work on self-tracking and conversational agents, for example, has explored how feedback and guided reflection can support transitions between work and non-work states or help users stick to goals [98, 99]. This study reinforces the value of reflection and highlights its role not just in optimizing existing strategies, but in re-engaging users who have disengaged.

3.9.3 Emerging strategies: task breakdown and manual time tracking

In addition to reflection, two strategies stood out as particularly helpful for participants trying to regain control over their schedules: breaking down tasks and manually tracking time. These strategies were not prompted or instructed, they emerged organically as participants attempted to improve their planning during the lockdown. Both strategies are notable because they echo findings from the planning bias literature: decomposing tasks into subtasks and drawing on past durations can improve estimation accuracy and reduce optimism [81, 69].

Breaking down tasks helped participants create more specific and manageable goals, which in turn made their plans feel more actionable. Rather than writing finish paper, they might write revise introduction or check related work section. This specificity helped reduce cognitive load and increased the likelihood of task completion. Similarly, manual time tracking allowed participants to compare how long tasks actually took with how long they thought they would take. In many cases, this prompted a recalibration of expectations and greater realism in future plans.

Interestingly, despite the availability of automated time-tracking tools, participants chose to track their time manually. They did so by noting down start and end times for tasks in notebooks or spreadsheets. This choice speaks to the reflective benefits of manual tracking. Unlike automatic systems that passively collect data in the background, manual tracking requires active attention and fosters awareness. Participants became more attuned to how they used their time and were more motivated to improve.

The preference for manual tracking also highlights several limitations of current tools. Many automated trackers monitor activity on apps or websites, but do not provide task-level context. They may be good at identifying distractions (e.g., time spent on social media) but poor at helping users evaluate how well they estimated or executed a particular work task. Furthermore, most do not support integration with planning data (e.g., goals or time estimates), making it difficult to reflect on plan accuracy. These gaps suggest several design opportunities: time trackers that include or link to to-do lists; features that ask users to rate or tag completed tasks; and dashboards that highlight discrepancies between planned and actual durations. Such tools could provide the benefits of manual tracking without the burden.

3.9.4 Designing for strategy, not just tasks

The core insight from this study is that supporting accurate planning is not just about helping people make better time estimates, it is about helping them maintain and adapt their planning routines. During periods of disruption, people may abandon planning entirely. When they do, they lose the scaffolding that supports focused work, prioritization, and time awareness. To rebuild that scaffolding, they need tools that do more than schedule tasks. They need tools that support the development, adjustment, and re-engagement of planning *strategies*.

For example, a system might prompt users to break down vague tasks, offer nudges to reflect on their day, or highlight discrepancies between expected and actual durations. More ambitiously, systems could model users' planning behaviour over time and suggest strategy changes when certain patterns emerge, such as frequent overestimation or failure to plan altogether. These systems would not just manage tasks; they would support planning as a skill that evolves over time.

This idea builds a bridge to the next chapters in this thesis, which focus more explicitly on strategy-based interventions for debiasing time estimation. While Study 2 shows that some users discovered helpful strategies on their own, it also suggests that many underestimated their value or did not know how to begin. This opens the door to exploring what kinds of tools and interventions can support the intentional adoption of planning strategies like task breakdown and time tracking. By understanding when people are most receptive to changeand what support they need to follow throughwe can design systems that help workers build planning skills that persist long after a disruption has passed.

3.9.5 Limitations specific to Study 2

It could be argued that having participants join the study at different weeks during the first month of lockdown has confounded the insights because people were going through different stages of adjustment. However, participants were asked to provide an overview of their experience week by week before and after the pandemic started. We observed consistency across participants in terms of how they felt the first weeks impacted work and planning, which indicates that this issue should not be of significant concern.

3.10 Summary of Chapter 3

This chapter examined the challenges of accurate task planning in academic knowledge work through two complementary studies. Study 1 combined diary and interview methods to capture everyday planning behaviour, showing that participants typically planned to work longer than they did and left about a third of their planned tasks unfinished. These inaccuracies were often linked to vague or incomplete plans rather than to external interruptions, suggesting that accuracy could be improved through stronger support for time estimation. Participants also described different approaches to planning, ranging from minimal to multi-level, indicating the need for tools that can adapt to diverse practices.

Study 2 extended the investigation by analysing weekly interviews during the first COVID-19 lockdown. The findings showed that when workloads and working conditions changed rapidly, many participants deprioritised planning in favour of immediate task execution, while others abandoned it altogether. Some participants, however, adopted new strategies that improved accuracy and control over time, most notably task breakdown and manual time tracking. Reflection played an important role in recognising the benefits of these strategies and motivating their continued use.

Taken together, the two studies demonstrate that planning accuracy is affected both by everyday cognitive biases and by contextual changes in work conditions. They also show that concrete strategies can improve planning accuracy but are not always easy for individuals to adopt or sustain. These insights motivate the shift in Chapter 4, which synthesises psychological evidence on debiasing to identify strategies that can inform the design of planning tools and evaluates the extent to which such strategies are supported in existing applications.

Chapter 4

Studies 3 and 4: Debiasing support in PTM tools

Parts of this chapter have been published in Ahmetoglu, Brumby and Cox [100].

Chapter Outline *This chapter presents two studies that examine how evidence-based debiasing strategies can be translated into design knowledge for planning technologies. Study 3 synthesises experimental literature to identify four strategies that improve time estimation accuracy: duration feedback, distributional data, task breakdown, and induced neutrality. Study 4 then evaluates the extent to which these strategies are implemented in practice through a functionality review of 47 personal task management (PTM) applications. Together, these studies highlight both the theoretical basis for supporting planning accuracy and the gaps that remain in existing technologies, laying the groundwork for the field interventions in Studies 5 and 6.*

This chapter begins the second phase of the thesis, which focuses on translating psychological theory into design strategies. Building on Studies 1 and 2, which demonstrated the prevalence of planning inaccuracies and the potential for improvement, Study 3 uses a rapid literature review to identify four strategies with empirical support for mitigating bias. Each strategy is examined in terms of its evidence base, mechanism of action, and relevance for integration into planning tools. Study 4 extends this work by investigating whether current technologies already support these strategies. Using a functionality review methodology, it analyses 47 widely used PTM applications. The findings show that only two of the four evidence-based strategies are implemented at all, and those are supported only infrequently or in limited ways. The other two strategies are absent, indicating a substantial gap between validated interventions and commercial tool design. By combining theoretical synthesis with systematic evaluation, this chapter develops a framework for understanding how planning accuracy can be supported in practice. The results clarify which strategies hold promise for integration into future tools, and which remain untested in real-world contexts. These insights inform the design of Studies 5 and 6, which examine how debiasing strategies can be operationalised and evaluated in field interventions.

4.1 Introduction to Study 3

Studies 1 and 2 established that academic knowledge workers in academia often plan tasks in ways that are vague and overly optimistic, leading to significant gaps between intended and actual work. While participants occasionally developed more accurate strategies such as task breakdown or manual time tracking these strategies were often discovered by chance, underused, or undervalued. This raises an important question for designers of planning technologies: to what extent do existing planning technologies support users in making accurate plans?

Although the digital marketplace is saturated with PTM tools, from simple daily planners to advanced task managers, there has been limited investigation into whether these tools actively support realistic time estimation. Prior HCI studies have explored user preferences and unmet needs in productivity tools through examination of current practice (e.g., [55, 26]), but few have bridged these insights with findings from psychological research on time estimation bias. Without incorporating strategies shown to reduce planning inaccuracies, PTM tools risk reinforcing rather than remedying optimistic planning.

To build a foundation for evaluating existing planning tools, this study conducts a rapid literature review (RLR) to identify evidence-based strategies that improve time estimation accuracy. Previous reviews have mapped causes of the planning fallacy [60, 59, 7], but have not translated this knowledge into actionable strategies for tool design. This study fills that gap by systematically identifying a set of strategies from psychology and behavioral science that are both theoretically grounded and practically implementable in PTM tools.

The findings of this review contribute four evidence-based strategies to support more accurate time estimations for everyday tasks. Synthesized from 28 experimental studies, these strategies include duration feedback, distributional data, task breakdown, and induced neutrality. Duration feedback and distributional data improve predictions by grounding them in prior task performance, whether through personal memory or reference to objective metrics. Task breakdown helps users identify specific steps within a task to make estimates more reliable, while induced neutrality reduces motivational pressures that can lead to overly optimistic forecasts. Together, these strategies provide a practical foundation for evaluating existing PTM tools in the next study, which examines whether and how these approaches are implemented in widely used applications.

4.2 Method of Study 3

This study employs a rapid literature review (RLR) methodology [101] to efficiently identify debiasing strategies for the optimistic planning bias, aligning with the goal of promptly informing the design of productivity tools. While systematic literature reviews offer comprehensive insights, RLRs enable the swift gathering of pertinent information while maintaining transparency and minimizing bias.

4.2.1 Search strategy

Studies were identified through a systematic database search in November 2023. An informal review of the literature indicated that most studies on this topic were published in APA journals. In addition, the ACM library was selected to ensure that studies in the field of HCI were not omitted. The searches were performed in the ACM Library: Full Text Collection and in Ovid: an interface that provides access to a range of databases including PsychNET. Search phrases were: planning fallacy, optimistic planning, estimation bias, predicted task duration and estimated task duration. Additionally, a "snowball search" was performed based on the studies obtained so far by searching for citations and by inspecting the reference lists of all available studies.

4.2.2 Eligibility criteria

As this study was interested in evidence-based strategies for accurate time estimations of individual daily tasks, this focus led to six inclusion criteria (i) studies with experimental design as opposed to case studies, reviews of the literature or ethnography studies (ii) peer-reviewed research studies or

Table 4.1: An adapted PRISMA flowchart [16] showing Study 1 selection process.

Identification	Number of records returned from the database search (n=643)	Number of records identified from other sources (n=12)
	Number of records remaining after removing duplicates (n=655)	
Screening	Number of records screened by title and abstract (n=655)	Number of records excluded (n=608)
Eligibility	Number of articles assessed by full-text (n=47)	Number of records excluded (n=19)
Included	Studies included in the review (n=28)	

studies described in doctoral dissertations, (iii) judgement-based predictions as opposed to tests of formal prediction models, as in large-scale construction projects models, (iv) studies manipulating a task-related factor such as breaking down a task into steps as opposed to a contextual factor such as different task deadlines, (see [60] p.253) or an individual differences factor such as trait procrastination [102], (v) factors related to correcting how predictions are made as opposed to correcting behaviours to match (optimistic) predictions (e.g. implementation intentions research).

4.2.3 Records identification

643 records were identified through the searches (n = 391 in the ACM Library and n = 252 in OVID). Additionally, 12 records were identified from the snowball search. The records were downloaded in Zotero and screened by titles and descriptions according to the inclusion criteria (n = 608 excluded). Next, the remaining 47 records were read in full (n = 19 excluded). As a result, 28 studies were included in the final review.

4.2.4 Data analysis

Data was extracted through a data extraction form with columns for factors that have been proposed to reduce the magnitude of the bias and success of the intervention. Synthesis was performed to distill practical strategies that can be used as advice on to how to manage everyday tasks. For example, *task breakdown* was identified as a debiasing strategy based on studies that explore factors such as *segmentation* [90], *unpacking* [69], *obstacles* [70]. These factors were grouped together because they all point to the importance of accurate identification of task elements. Research group discussions were held during the synthesis process where I presented results about the strategies and subsequently performed several iterations based on feedback received to arrive at the final set of strategies.

4.3 Results and Discussion of Study 3

The findings show four strategies that PTM apps should encourage to support accurate time estimations for daily tasks. The first two strategies are based on the principle that information about the past should be used to inform predictions about the future: (i) duration feedback, the regular and objective feedback on duration of recently completed relevant tasks which improves and promotes use of memory when planning; (ii) distributional data, the reliance on past data about duration of similar tasks as the foundation for predictions. The other two strategies identified here are focused on improving the process of planning itself to avoid common pitfalls: (iii) task breakdown, the breaking down of tasks into steps and estimating duration for each step which allows for correct identification and estimation of more specific task components, and (iv) induced neutrality, the use of a neutral perspective to time estimations which avoids strong motivation to quickly complete tasks and elongates biased estimates.

4.3.1 Feedback

Thomas et al. [103, 104] ask participants to estimate the duration of a just completed task before making a prediction. In a series of studies, they show that subsequent task predictions are anchored on the recalled duration of just completed tasks. These results mean that self-generated feedback influences time predictions. Konig et al.[105] further demonstrate that when participants are asked to estimate the duration of tasks completed a week ago, they become more accurate at estimating the duration of a future task compared to when they are not asked to give retrospective estimates.

Brady et al. [106], explore the effect of a set of reflective questions about past completion times on future predictions. They find that asking participants about how difficult they have found performing similar tasks in the past is not sufficient to reduce the planning fallacy. They suggest that the intervention did not produce an effect because participants mostly gave answers only few words long. Buehler et al. [61] demonstrate that asking participants to provide longer answers about how past experiences with similar tasks are relevant for their future plans mitigated the bias.

Roy et al.[107] found that participants performed better when objective measures about past performance is provided to them compared to reliance on memory. Tobin and Grondin [71] demonstrate that expert runners are more accurate at predicting and estimating the duration of 5k running competitions compared to intermediate runners. They argue for effect of prior experience due to extensive feedback received during the training of experts. Effects of feedback during training interventions however have not showed promising results so far. Gruschke and Jørgensen [108] and Jørgensen and Gruschke [109] create practical sessions where software developers learn how to increase skills in estimation however the studies did not find a decrease in estimation accuracy for software projects.

These studies suggest that effective feedback on task duration needs to be timely, relevant, objective, and extensive. These findings build upon previous reviews of the literature. Buehler and Griffin [7] point out that drawing on past experiences reduces the magnitude of the bias. Roy emphasizes the utility of objective feedback [59]. Halkjelsvik and Jørgensen [60] note that translating the effects of feedback to practical training in software development has had limited success. However, while feedback has been suggested as one factor for improving estimation accuracy, previous research has not provided clear principles about what the feedback should entail. Building upon these findings, the present review outlines key principles for effective feedback implementation.

4.3.2 Distributional data

Distributional data involves gathering data about a set of past similar tasks to support predictions for time. In an experiment by Roy [107], participants received average tasks duration before predicting the time required for various tasks, such as receiving a food order. Those with access to distributional data provided more accurate estimates compared to a control group. Similarly, Shmueli et al. [110] provided undergraduate engineering students with reference information about past completion times for software projects. Students with this information estimated the project duration to be 28 hours, whereas those without estimated it to be 18 hours.

Friedorf and Buehler [111] tested *reference class forecasting*, a standardized procedure commonly used for predicting large-scale project duration [112, 113]. Participants in their studies followed steps to recall similar projects and derive an average estimate for future duration. This procedure effectively mitigated bias in three experiments. Similarly, Koval and Jansen [73] studied the effect of predictive visualizations constructed from participants own elicited beliefs about a single task. In their crowdsourced experiment, participants decomposed an imagined grocery-shopping trip into sub-tasks and possible surprise events, provided prediction intervals and likelihoods for each, and the authors used these inputs to simulate a personal mixture distribution of plausible durations. Visualizing this distribution as a quantile dot plot led participants to increase both their point estimates and interval widths, indicating reduced optimism in their duration estimates.

Existing efforts in HCI for automatic duration estimation [114] suggest that it is feasible to predict tasks planned duration based on attributes such as task name and creation/completion times using machine learning algorithms. However, it is not clear how these estimates relate to the actual time people take to complete planned tasks. It also remains unclear whether current apps employ any approaches for automatic duration inference, and how accurate these are. Investigating this aspect forms one of the key objectives of the next Study 4.

4.3.3 Task breakdown

Kruger and Evans [69] asked participants to estimate how long it would take them to complete one of several tasks: holiday shopping, getting ready for a date, formatting a document and preparing food. Those participants who were asked to break down the task before making a duration estimate provided longer and less biased estimates. The effect was more pronounced for more complex, multi-step tasks, a hypothesis also confirmed by Hadjichristidis et al. [115].

Forsyth and Burt [90] asked participants in four experiments to allocate time to single tasks or to components of those tasks separately. They found that the summed duration estimates for the components that made each task were overestimated meaning that the planning fallacy was reversed. This suggests that smaller tasks are easier to estimate. In addition, Siddiqui et al. [116] showed that simply asking people questions starting with *how* they would do a task compared to *why* minimized the underestimation bias for longer, more complex tasks because it made people more aware of the steps involved. Finally, Wiese et al. [117] demonstrated that working out steps of tasks back from a deadline produced more accurate plans because it prompted people to think about possible delays and obstacles compared to planning in the usual forward way. Interestingly, asking people to explicitly make a step-by-step plan can also make them more optimistic [118], an effect explained by study instructions that encourage people to focus on the future and plans for success, considering

Table 4.2: Summary of Study 3 results: strategies for increasing accuracy of time estimations for everyday tasks

Strategy	Summary	Citations
Feedback	Regular and objective feedback on duration of recently completed relevant tasks supports and promotes use of memory in time estimations.	[103] [104] [105] [106] [61] [107] [71] [108] [109]
Distributional data	Relying on past data about the duration of similar tasks as a foundation for time estimates bypasses human judgement and improves accuracy.	[107] [110] [111] [73]
Task breakdown	Breaking down tasks into smaller tasks and estimating the duration for each step separately supports easier and accurate time estimations.	[69] [115] [90] [116] [117] [70] [119] [70, 119]
Induced neutrality	Employing a neutral perspective in time estimations reduces motivation to quickly complete tasks and elongates biased estimates.	[121] [122] [123] [61] [124] [110] [125]

the obvious steps within a task, and omitting the non-obvious ones.

Researchers have additionally explored if people can adopt a more pessimistic perspective to counter for optimistic time predictions by asking participants to list potential obstacles, delays, and interruptions, and to imagine pessimistic scenarios [70, 119]. However, these attempts have yielded largely unsuccessful results. The lack of effectiveness in these interventions may be attributed to a reluctance to change one's mind when confronted directly with disproving evidence. Individuals struggle to accept the possibility that plans will most likely fail as that would cause a cognitive dissonance[120].

These findings show that allowing people to discover themselves possible delays and obstacles through the process of task breakdown is one promising strategy. It is also facilitated by easier estimations for smaller tasks. Therefore, the task breakdown strategy recommends for a detailed task decomposition focusing on both obvious and non-obvious steps within tasks, and advises exercising caution against direct instruction to consider pessimistic scenarios or obstacles. These findings build upon prior reviews by synthesizing different factors (decomposition, focus on obstacles, pessimistic scenario generation [7, 73, 60]) into a practical guideline.

4.3.4 Induced neutrality

Most strategies above explore a cognitive factor for improving time estimates. Researchers have additionally explored the potential of motivational factors such as task involvement and incentives. Henry and Sniezek [121] show that individuals overestimated their future performance the least when perceived internal control was low and when no monetary rewards were involved. Weick and Guinote [122] suggest that reducing people's sense of power over the future leads to less optimistic predictions. Lohre and Teigen [123] show that simply question wordings that take away the focus on the individual and focus on the task unfolding instead reduce the underestimation bias. They compared asking *how fast can you do* a task with *how long will it take* and found that the latter resulted in less bias. These studies in essence reduce one's sense of involvement with the task in advocate for neutral attitude in estimations.

Research also has induced neutrality by instructing people to adopt a neutral observer perspective. Buehler et al. [61] found that prediction times for academic and non-academic times were overestimated (the bias was reversed) when observers made predictions for actors however Roy [125] did not find the same effect. To explore further the potential use of observers, Buehler tested the effect of imagined observer perspective. In four studies, Buehler et al. [124] tested the effect of inducing an observer perspective with task instructions (third-personal imaginary). They asked participants to identify a specific upcoming task, to imagine it unfolding and predict when they would complete it. Participants in all four studies predicted longer completion times when they imagined an upcoming task from the third-person rather than first-person perspective. Third-person imagery increased focus on potential obstacles and decreased the impact of task-relevant motives on prediction. These findings were replicated in the realm of software project estimation by Shmueli et al. [110].

4.4 Summary of Study 3 and bridge to Study 4

Study 3 addressed the second research sub-question of this thesis: *To what extent do existing planning technologies support users in making accurate plans?* In order to lay the foundation for evaluating planning tools, this study conducted a literature review to identify evidence-based strategies for improving time estimation accuracy in everyday tasks. Twenty-eight experimental studies were analyzed, and four main strategies were synthesized to inform the design of personal task management applications. First, the use of feedback about the actual duration of past tasks, especially when it is timely, objective, and detailed, was suggested to significantly improve future estimates. Second, reliance on distributional data: information about the duration of similar past tasks, was suggested to reduce underestimation and support more realistic planning. Third, task breakdown, or segmenting a task into smaller actionable steps, helped users uncover hidden complexities and provided a more reliable basis for estimation. Finally, induced neutrality, achieved by reducing personal involvement, e.g. by adopting an observer perspective, mitigated motivational biases and led to longer, more realistic predictions. This synthesis provides a practical foundation for the next study, which examines whether current PTM tools incorporate these evidence-based strategies and explores opportunities for better supporting users in managing their tasks.

4.5 Introduction to Study 4

This study addresses the second subquestion of the thesis: *To what extent do existing planning tools support strategies that improve time estimation accuracy?* While the previous chapters established the prevalence of planning inaccuracies (Study 1), showed how workers adapted their routines in response to disruption (Study 2), and identified four evidence-based debiasing strategies (Study 3), this chapter examines how well these strategies are currently supported in existing personal task management (PTM) applications. It evaluates whether current tools align with psychological theory and identifies gaps that can inform the design of future interventions.

The market is proliferating with PTM software applications, such as Todoist, Google Tasks, Microsoft ToDo, Trello, Basecamp, to name a few. These apps provide a digital substitute to the traditional paper counterparts. Studies show that users often combine digital and non-digital tools for time and task management, and the reported rates of digital apps adoption did not match the initial expectations [2]. The slower-than-anticipated adoption rate can be attributed to a significant number of users abandoning digital tools, either reverting to traditional paper-based methods or switching to alternative tools due to unfulfilled user needs [54]. Studies show strong user preferences for tools that are flexible, informal, natural, personalizable and intuitive. HCI studies have taken these needs in consideration when designing PTM apps prototypes that resemble text editors and feel more natural (e.g. Plan, [117]) or can be customized to meet individual needs (e.g. Scriper, [55]). While this direct digitalization approach has many strengths for improving user friendliness and adoption rates, it does not allow for design to account for existing cognitive biases such as the planning fallacy. Regardless of the tool type employed, the planning fallacy persists unless the design incorporates interventions aimed at mitigating it.

PTM apps are an obvious candidate for embedding interventions targeted at tackling the planning fallacy. While users report a need for more planning support from their digital tools (incl. in Studies 1 and 2) [77, 78, 2], there are a limited number of HCI studies that test interventions for improving planning skills that can inform the design of apps. For example, Williams et al. [99] designed a conversational agent, SwitchBot, that supported users to psychologically detach from work by planning and reviewing daily task lists. They found that people felt more productive at work and sent fewer after-work emails while using the bot. In addition, when prompted to reflect on their work habits, people report identifying new goals related to better planning skills [27] in the interviews in Study 2 which prompted participants to reflect and come up with new strategies. However, these studies explore the benefits of planning in a general sense instead of discussing specific HCI interventions for improving planning skills and reducing the planning fallacy.

With the proliferation of digital applications and the expanding user base, there's a growing need in evaluating the effectiveness of these tools in meeting user needs within the field of HCI. To address this need, researchers are increasingly turning to methodologies such as functionality reviews to scrutinize the landscape of popular digital applications through the lens of user needs or theory-informed frameworks. Stawarz et al. [126] aimed to understand the type of cues that facilitate habit formation and apply those insights to a review of 115 habit formation apps. The first study in the paper found that relying on reminders supported repetition but hindered habit development, while the use of event-based cues led to increased automaticity. The functionality review revealed that existing apps focus on selftracking and reminders, neglecting support for event-based cues. The authors advocate that apps have significant potential to offer genuine habit support and they propose design guidelines aimed at facilitating habit formation through the incorporation of contextual cues. In another functionality review study inspired by Stawarz et al. [127], Lyngs et al. [128] reviewed 367 apps and browser extensions aimed at improving focus and reducing distractions while interacting with devices. The authors employ a dual systems model of self-regulation and apply it as a framework for organizing and evaluating design functionalities. Through this analysis, the paper contributes by understanding how existing tools align with or differ from established self-regulation models, and also by revealing underutilized cognitive mechanisms to guide the design of new functionalities.

Similar functionality review studies have been conducted with stress management apps using personal informatics and behaviour change theories [129], health and wellness apps and assessing their behaviour change potential [130], systematic descriptions of preconception care apps and analysis of their user reviews [131], review of the extent to which gamification is used in fitness apps as a health behaviour factor [132] and others. Despite this trend, there has been a notable absence of comprehensive reviews specifically focusing on PTM apps. Bridging the gap between theory and practical application, the current paper seeks to address it by thoroughly reviewing PTM applications using a research-informed framework.

4.6 Method of Study 4

This functionality review study has three steps: first, popular PTM apps are systematically identified, and their functionalities are fully examined by three researchers. Next, a mapping between the strategies from Study 3 and functionalities is performed to narrow down those that could support the strategies. Finally, an evaluation is conducted by revisiting the apps. During this phase, observations are recorded about the steps required to achieve strategies with respective functionalities, ease of use is assessed in relation to the desired strategies, and the alignment between the strategy and functionality is confirmed.

4.6.1 Search strategy

Web articles featuring rankings and recommendations of PTM tools were selected as the source for identifying applications. Typically, studies reviewing applications utilize app stores as their primary data source. However, in the context of work-related desktop applications, app stores proved less useful due to the vast number of available apps, many of which were mobile-oriented and not widely used for work purposes. Moreover, many basic PTM apps without significant user bases clutter app stores, making it impractical to thoroughly review the numerous available options. Therefore, the decision to utilize web articles reflects a desire to focus on a smaller, more relevant subset of applications and to provide a more realistic representation of how individuals might search for task management tools for work.

Web articles with the terms of *best to-do list* and *best daily planner* were searched in April 2022 through two search engines: Google (n=27 articles) and DuckDuckGo (n=28 articles) which con-

Table 4.3: An adapted PRISMA flowchart [16] showing the app selection process in Study 4.

Identification	Apps identified from DuckDuckGo and Google web articles (n=283)	Duplicate web articles removed (n=21) Duplicate apps removed (n=182)
Screening	Apps screened by descriptions (n=101)	Number of records excluded (n=45)
Eligibility	Apps downloaded and assessed (n=56)	Number of records excluded (n=9)
Included	Apps included in review (n=47)	

tained a varying number of PTM app recommendations: a total of 283 apps. DuckDuckGo was chosen alongside Google because it does not tailor search results based on user search history or personal preferences, ensuring a more neutral and unbiased search environment. After removing duplicates, there were 21 articles containing 101 unique apps.

4.6.2 Eligibility criteria and apps identification

The identified 101 PTM apps were screened by descriptions according to inclusion criteria: main use for creating to-do lists (15 excluded), a PC software application (12 excluded), aimed for use in individual planning context as opposed to team planning only or having a plan for individual use alongside team plans (11 excluded), in English (1 excluded), being able to access the app (2 apps with technical faults and 2 discontinued). The remaining 56 apps were screened in full by signing up for the app either with a free trial or paid subscription if free trial was not possible. Apps were opened in a Chrome browser on Windows 10 Enterprise HP Elitebook laptop, or downloaded and installed on the same laptop if only a desktop app was available. Forty-seven apps were included in the final review, with 5 excluded due to other main purpose, 3 were not for individual use and 1 had a technical fault.

4.6.3 Functionalities identification

A functionality was defined as *the set of actions that the app offers to help users manage their tasks and activities efficiently.* This broad definition encompassed all functionalities offered by apps related to task management. A codebook was developed after reviewing 20 randomly selected apps. Then, the codebook was applied to the rest of the apps with revisions along the way. Subsequently, the codebook was re-applied to the apps that were missing a rating on the recently identified functionalities. This coding process was conducted by the first author in May and June 2022. Reliability of the codes was ensured by asking two additional researchers to apply the codebook to ten randomly selected apps. Cohen's Kappa was calculated and indicated substantial agreement (.78). A total of 188 functionalities were identified across the 47 apps (See Appendix A.1 and Appendix B.1). Each functionality was marked as either *yes* or *no* for each app to denote its presence. An average prevalence was calculated based on these data to determine the occurrence of each functionality across the apps.

Mapping between strategies and functionalities

The corpus of functionalities was used to identify those that could support the use of the strategies identified in Study 3. Initially, I attempted to relate all identified functionalities to the strategies in a direct way by mapping each functionality with its definition from the codebook against each strategy. This mapping attempt was done during research team meetings where the I provided a list of all functionalities, their definitions and a descriptions of the strategies from Study 3.

Despite detailed descriptions and definitions, the codebook did not adequately illustrate how each functionality operated in relation to the strategies, requiring raters to possess prior experience with the PTM app. The research members of the team were unsure about the actions each functionality allowed users to do and found the task challenging. Consequently, these group sessions were utilized to refine the potential functionalities for implementing each strategy. A list of 11 functionalities were narrowed down from the corpus as potential candidates for allowing the use of the strategies from Study 3. Researchers shared and deliberated their opinions, after which I revisited the apps in December 2023 to perform an evaluation of the narrowed down list of functionalities.

I re-visited at least half of the apps that offered the respective functionality and spent up to 30 minutes attempting to implement it while taking notes with observations. Observations were made about: steps involved in implementing the strategy, confirmation of the fit between the functionality and the strategy, and observations about ease-of-use of the functionalities to achieve the aims. At that stage, the functionalities were further narrowed down to 5. I then made a final set of observations by engaging with the functionalities for a longer duration of time. Typical academic tasks, such writing a research proposal, were used as examples and I tried to apply the strategies through the functionalities over the course of at least one day. The final set of functionalities and the respective observations are presented in the Results section below.

4.7 Results of Study 4

Study 3 identified four strategies for improving the accuracy of time estimates: duration feedback, distributional data, task breakdown and induced neutrality. Findings show that three functionalities can support users to implement the first strategy: time tracking, pomodoro session and time analytics; no functionalities could support users to implement the second strategy, even though we identified custom analytics as one potential candidate but our attempts were unsuccessful; and two functionalities: subtasks and templates could support users in implementing the third strategy of task breakdown. No functionalities were found to support the fourth strategy of induced neutrality. We now turn to discussing each functionality separately, describing its nature, prevalence, alignment with the strategies and its ease of use in this specific context (see table 4.4). Recommendations are made for each one to improve the likelihood of users engaging with them according to the strategies.

4.7.1 Time tracking

Tracking time spent on a task allows user to record time data in order to obtain feedback. A minority of apps (n = 11, 23%), supported time tracking as a native functionality. Out of those, 7 apps enabled this functionality by default (e.g. Wrike and Jira) and 4 apps (e.g. Asana and ClickUp) required users to enable the functionality from settings or custom fields if they were interested in using it. Almost half of the apps in this study (n = 20, 43%) could be customized in a way to support time tracking either through native integration with another app or through a browser extension (e.g. Todoist and Time Doctor, Zenkit and Clockify, TMetric and Microsoft ToDo) or through a custom integration with Zapier (e.g. Lunatask and Timely Time Tracking). A third of apps (n = 16, 34%) did not support time tracking and could not be customized for time tracking (e.g. Habitica and TeuxDeux).

To begin tracking time, users must first define a task. This involves selecting an appropriate description that will aid in the time tracking process. Once defined, users can then utilize either an automatic timer or manually input the duration for each task, depending on the app's functionalities. Typically, most apps offer both options (such as the one in Asana illustrated in figure 4.1). When opting for the timer, users need to return to the app to manually initiate and conclude timings as they transition between tasks. The accumulated time spent on a task can then be reviewed within the time tracking menu associated with each task.

The process of locating and using time tracking was straightforward. We were able to find the time tracking functionality and launch it without any confusion. Using both the timer and the manual log option was quick and easy and produced easy to understand time logs. Those logs could easily be adjusted at any point and we could track times for tasks in precise units and also tasks spanning several sessions of work which were often annotated with user identifiers and log lists.

Despite these advantages, we noted several challenges that became clear when transitioning from tracking individual task duration to employing time tracking over extended periods, such as a full day. First, existing time tracking implementations require high user motivation and previous knowledge. For example, defining a task a user is interested in tracking poses difficulties. Most users do not have experience with time tracking or detailed daily planning. They may not necessarily know what the best way to define their *tasks* is, and may lack clarity about the tasks they want to do during the day. Deciding on what they wish to do and actually sticking to this plan can be perceived as burdensome [77]. In addition, time tracking often requires users to remember to start and stop timers or log their time at regular intervals. Therefore, recording time for the smaller tasks, which is a better fit with the strategy, may lead to a more time-consuming process prone to errors. The constant attention to tracking time can be distracting and disrupt the flow of work if users decide to track small tasks.

Second, time tracking functionalities may not be discovered by those who need them. In many cases, especially for project management-focused apps, task time tracking was primarily geared towards recording billable/contracted hours leading to complicated time tracking features such as fields for hourly fees. A very limited number of apps (n = 3, 6%) offered native time tracking functionalities aimed at individuals looking to support personal productivity. Time tracking was only then explicitly implemented as a tool for enhancing personal outcomes. When looking to support personal productivity, the majority of other apps often directed users to automated tracking of the number of *tasks* completed during the day, instead of tracking *time* for tasks (e.g. Todoist's *Karma*). While this may allow users to approximate the amount of work that can done within a day, it is a not fully in line with the strategy of duration feedback.

These findings point to the need for more learning and coaching support in PTM apps. To mitigate the observed challenges, the app could guide users to focus on targeting specific user tasks, particularly those prone to bias (see [77]), or encourage tracking during periods of heightened user motivation (see [54]). Moreover, if a user begins to disengage from tracking, the app could suggest grouping tasks together and tracking time under broader categories such as *writing* or *deep work* to

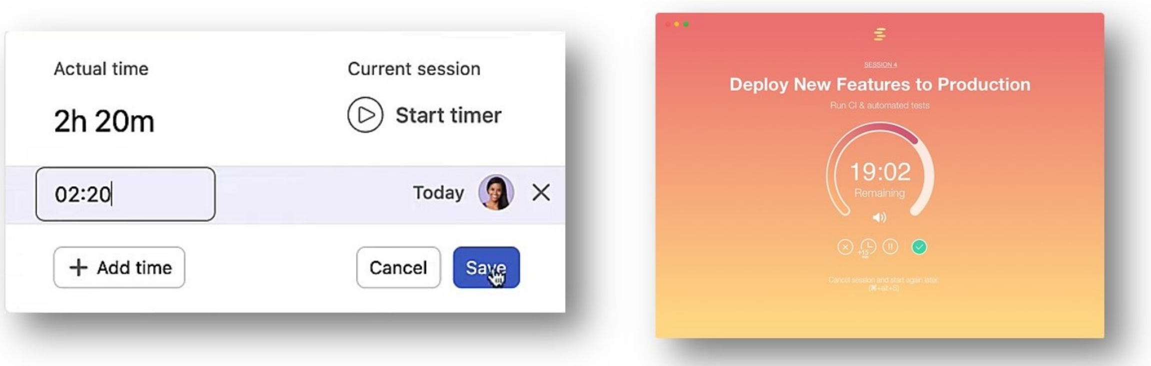


Figure 4.1: An illustration of time tracking in Asana (left) and pomodoro session in Serene (right)

reignite user interest in more detailed tracking. Moreover, apps can promote time tracking by prompting users to log their activities or time spent on tasks. For example, Super Productivity app prompts users to label time spent away from the desktop upon return. Another example is Sunsama, an app that includes *rituals*, guided step-by-step planning routines to support users in setting, tracking and reflecting on tasks. However, proactive time tracking systems should be carefully implemented to avoid being perceived as unrealistic. For instance, Super Productivity sends alerts such as *Take a break! You've been working for 72h* if the user does not stop the timer, rendering the approach ineffective.

However, due to the low number of apps that currently promote time tracking to individuals, the development of those features may lag behind other more user-driven features, such as focus-aiding and smart reminding. This piece of research puts forward that by tackling the root causes of user challenges, we could reduce the need to implement short-term interventions aimed at stress reductions due to a feeling overwhelmed with tasks and deadlines.

Automatic time tracking is one way to reduce user burden of tracking that has been explored in previous HCI research [133] and purpose-built tools. Examples of automatic time tracking software include RescueTime and ManicTime. These tools track users' computer usage and categorize their activities automatically. However, none of the current apps included an obvious and easy integration with automatic time tracking software. A future challenge lies in integrating them seamlessly with PTM apps [9]. Research suggests that a combined approach of manual and automatic tracking (semi-automatic) may work best to foster awareness [134] and external rewards may be needed to encourage users to engage with automatic tracking [135].

Table 4.4: Summary of Study 4 results. It shows the strategies for improving time estimation accuracy identified in Study 3 together with the functionalities that support users in implementing them and recommendations for designers of PTM software

Strategies	Functionalities	Recommendations for designers
Feedback	Time tracking	Explore semi-automatic and proactive time tracking to reduce user burden.
	Pomodoro session	Support multi-user pomodoro sessions.
	Time analytics	Provide options for regular reviews of time analytics data on individual level.
Distributional data	-	Provide time analytics for easy aggregation of similar tasks' duration over time.
Task breakdown	Subtasks	Alert users when tasks are too vaguely defined and encourage them to break them down.
	Templates	Provide personalised templates.
Induced neutrality	-	Incorporate dialogues such as conversational agents.

4.7.2 Pomodoro session

A Pomodoro session is a functionality that allows users to track time for different tasks in intervals typically set for 25 minutes, known as "Pomodoros," followed by a short break of around 5 minutes. After completing a certain number of Pomodoros, usually four, users are encouraged to take a longer break of around 30 minutes. The timer within the app notifies users when each Pomodoro session begins and ends, providing feedback on how much time has passed while working on a particular task. Pomodoro timers were included in a small number of apps in the sample (4 apps, 9% of apps). It offers features such as customizable work and break duration and task tracking within each Pomodoro session.

The Pomodoro timer presents an alternative approach to traditional time tracking, offering several advantages that address the challenges identified in the previous section. First, unlike traditional time tracking, where users may diverge into multitasking and be unsure about how to log time after doing several things at once, the Pomodoro timer encourages workers to focus on one small step at a time. In this respect, the Pomodoro timer also fits with the task breakdown strategy. Further, the Pomodoro timer may reduce the burden of constant user input compared to traditional time tracking. The Pomodoro timer operates automatically in the background, initiating and ending each Pomodoro session without requiring user intervention. This automated process alleviates the need for constant attention to time tracking, allowing users to concentrate on tasks and reducing the perceived burden of tracking time. Lastly, the close integration of the Pomodoro technique into the planning process could accelerate improvements in time estimation accuracy. Users set specific tasks for each Pomodoro session which allows them to receive real-time feedback on their productivity and task duration.

While previous research has acknowledged the benefits of the Pomodoro technique related to focus and task completion, this study suggests that another unexplored benefit may be enhanced time estimation accuracy [136]. However, it's important to recognize that following a strict predetermined work and break schedule may not be suitable for every environment or individual's preferences. Some workers may benefit from additional structure and external rewards, such as in collaborative sessions where co-workers meet virtually or in person to synchronize their work patterns. Previous research has observed benefits of such collaborative Pomodoro sessions for academic writing tasks, indicating potential avenues for further exploration [137].

4.7.3 Time analytics

Time tracking and Pomodoro sessions allow recording of task duration whereas time analytics present these data back to the user, and used together they can support the strategy of feedback identified in Study 3. Time analytics encompasses the process of analyzing tracked time or time spent in Pomodoro sessions, offering users insights through textual summaries or data visualizations. Within the scope of this study, time analytics were identified in a quarter of the apps (n = 12, 26%).

Study 3 emphasised the importance of timely, objective, relevant and extensive duration feedback for increasing the accuracy of time estimations. Time analytics align closely with these guidelines, as they provide an objective overview of where time was spent. Furthermore, in contexts where time tracking is pursued for personal productivity gains rather than billing purposes, immediate rewards may be lacking, potentially leading to a diminished motivation to sustain the tracking habit [138]. Time analytics in PTM apps can bridge this gap by providing users with immediate feedback and insights into task duration, thus supporting the habit of task tracking, fostering a sense of accomplishment and improving the accuracy of time estimations [78].

Despite the benefits associated with time analytics, certain limitations exist within current implementations. Most applications primarily focus on providing analytics at a team or organizational level

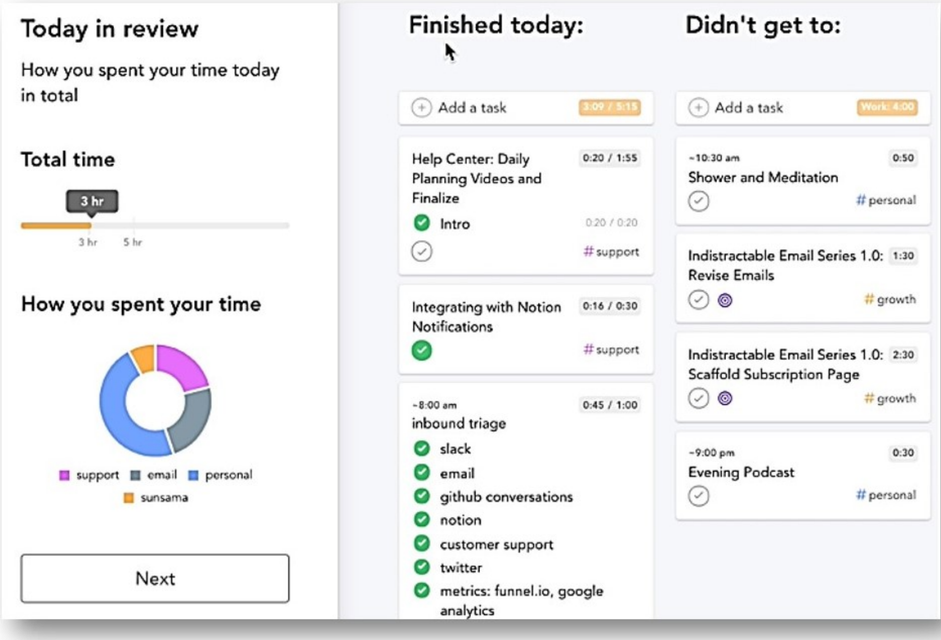


Figure 4.2: An illustration of time analytics in Sunsama's daily review ritual

or over large time scales (e.g. Teamwork). While such analytics may be valuable for managerial or strategic purposes, they may not cater adequately to individual users seeking to improve their productivity and time management skills. Consequently, similar to time tracking, fewer apps are tailored toward offering time analytics that cater to the needs of individual users. A small number of apps (n = 4, 9%) adopted this approach and proactively supported users in tracking and reflecting on time analytics data. Sunsama's rituals (see figure 4.2) and Super Productivity's "finish day" feature encouraged users to reflect on how time was spent during the day. Supporting and encouraging engagement with the data obtained with time tracking would further contribute to the success of this functionality.

4.7.4 Custom analytics

Time analytics did not support any default ways to aggregate distributional data in the provided charts, graphs and textual summaries. To explore the extent to which apps support this guideline identified in Study 3, we turned to the functionality of custom analytics. Custom analytics enable users to tailor the design and parameters of charts, graphs, or visual dashboards, providing them with an overview of their time and task management statistics.

The findings showed that custom analytics had a prevalence of 17%, present in 8 out of the 47

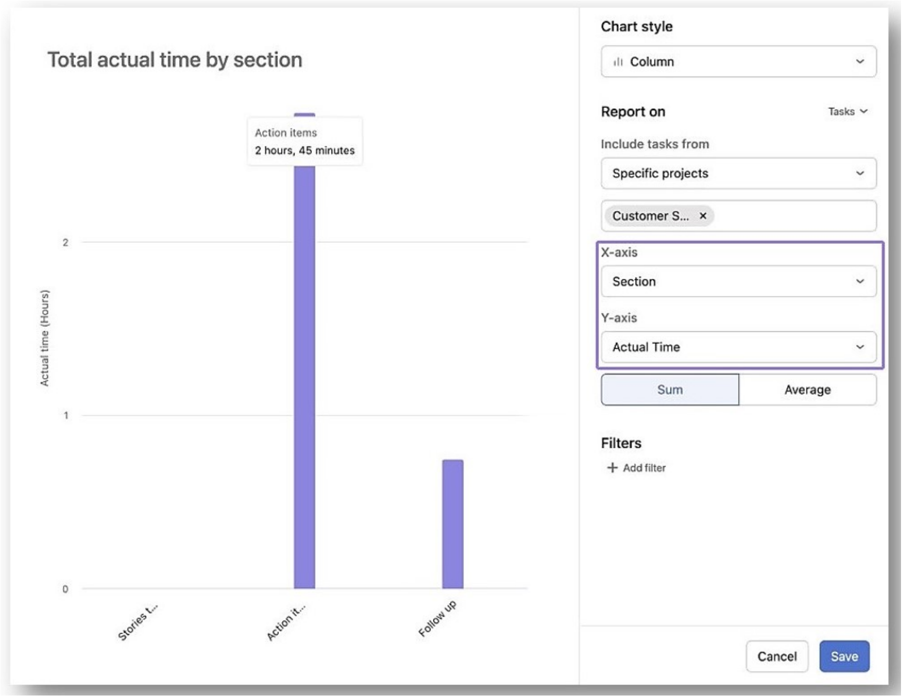


Figure 4.3: An illustration of custom analytics in Asana

apps reviewed. This functionality tends to be supported by feature-heavy PTM apps geared towards team and project management activities (e.g. ClickUp, Asana, Wrike). Such apps often cater to the needs of managers who may require flexible data visualizations for handling complex projects. Given the definition and nature of this advanced functionality, we expected that it would support users implement the strategy of distributional data.

There was no evidence found that custom charts effectively supported users in implementing the strategy of distributional data analysis. First, the efforts to visualize data concerning various types of tasks over time highlighted the necessity of labeling common task types. This labeling was essential to enable the chart builder to discern and categorize the data accurately. However, there was an absence of guidelines about how to label data so that the chart could then pick up correctly on the categorization. This led me to consider many different ways of labelling in attempt to guess the right answer. The chart-building process primarily relied on drop-down menus, which posed challenges in comprehending the implications of each option (see figure 4.3). It was not possible to sort the time tracking data according to the labels it had.

Despite concerted efforts, constructing the desired visualizations proved time-consuming, with none completed as an outcome. However, it is possible that there is a way to customize time analytics that was beyond my discovery. Therefore, I decide to keep the observations about this functionality and present the findings as an unsuccessful implementation of the strategy. A way forward may be including natural language inputs to articulate data requirements effectively. Moreover, even if the desired charts and graphs had been successfully created, they would have been nested within a separate analytics environment, disconnected from the task planning process occurring within the core task management areas of the apps. Nesting such charts within the task environments would make them useful for supporting time estimations.

4.7.5 Subtasks

Subtasks in PTM apps allow users to break down tasks into smaller tasks. Subtasks were a widespread functionality. Simple subtasks are often referred to as checklists when subtasks can be simply ticked off (n = 36, 77%). In other cases, subtasks can be assigned attributes such as due date, notes or priority (n = 29, 62%). To create subtasks, users begin by selecting or creating the main task. Usually located in a task menu, subtasks can be added. Multiple subtasks can be created and most apps allow the order to be rearranged. Some apps, such as Trello, allow progress tracking in the main task based on number of subtasks completed.

Subtasks per se may not necessarily lead to reduction of the planning fallacy. The fit between subtasks and the strategy of breakdown depends on how subtasks are used. If users focus on step-by-step plans and obvious subtasks, subtasks can increase the optimism of plans [118]. If users engage in more deliberate task breakdown and think about non-obvious steps, they may achieve more realistic plans.

Another limitation of subtasks related to ease-of-use was that subtasks were often tucked away within task menus, rendering them less visible to users. This observation suggests a potential underutilization of this functionality, a phenomenon corroborated by cognitive psychology research positing that individuals do not inherently decompose their tasks [61], and HCI studies indicating that workers tend to formulate loosely defined plans [77]. Consequently, adopting a more proactive approach to encourage task breakdown could prove advantageous for users and enhance the accuracy of time estimations. For instance, applications could deploy alerts when tasks are overly ambiguous, prompting users to further decompose them. Alternatively, users could be incentivized to create tasks with duration not exceeding 30 minutes, fostering a more granular and manageable task structure.

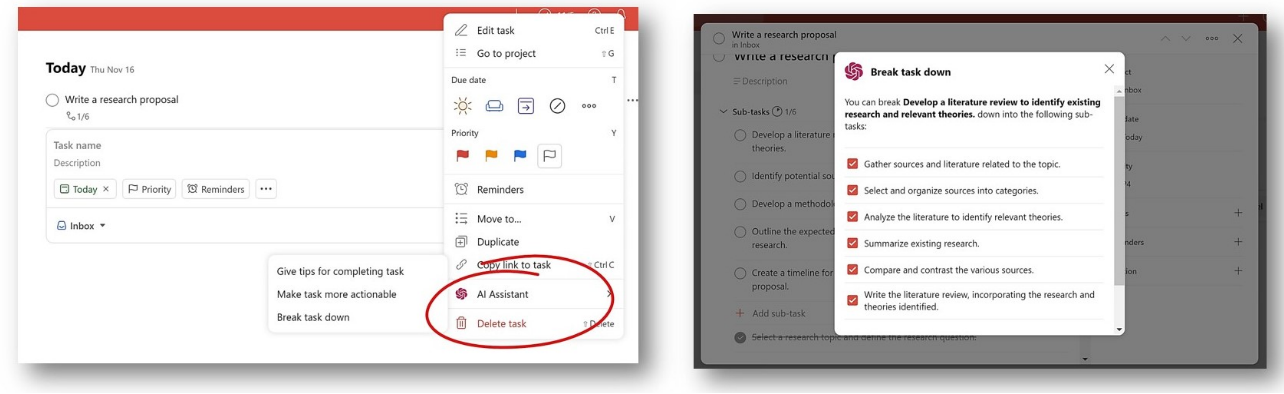


Figure 4.4: An illustration of Todoist AI assistant and auto-generated subtasks suggestions

4.7.6 Templates

Templates are ready-made task lists that have a specific purpose, such as a weekly review template or 1-on-1 meeting template. Templates were supported by 40% of apps (19/47 apps). They are sometimes located in their own database within the app and other times users can pick one when creating a new list. Also, some apps offer them upon signing up. Templates are specific task and projects breakdown that can support the strategy of task breakdown by providing the steps necessary for task completion. Users are able to create templates or borrow ones from a directory. Todoist, for example, is one of the most popular PTM apps on the market. It has a centralized directory of templates created and evaluated by the app team.

While templates may support task breakdown, the success of the strategy will depend on the quality of the template. One limitation is that templates were limited number for the most common tasks and are not applicable to most tasks workers do on everyday basis. To resolve this, apps such as Todoist have started to harness AI generative features to support a case-by-case generated subtasks (enabled through the AI assistant feature). The findings showed that while this innovation may be useful for brainstorming ideas for subtasks, this feature could be perceived as too generic failing to provide personalised recommendations for task breakdowns. For example, the template was not informed about how much experience the worker had with the task or other characteristics that may influence the choice of subtasks.

4.7.7 No functionalities identified for the strategy of Induced Neutrality

No functionalities were identified within the corpus to support the strategy of induced neutrality. The discrepancy may be attributed to the absence of dialogue features and conversational interactions, coupled with the reliance on form-based interfaces in PTM apps. Previous research suggests that workers may derive benefits from dialogue-based features for planning [99, 139], and their integration into PTM apps with the aim of induced neutrality could potentially enhance their effectiveness. For example, the aforementioned *Rituals* feature in Sunsama includes questions such as listing *obstacles in my way* during the daily planning ritual. According to Study 3. this approach may not be the most effective and a more neutral way of reducing optimism is of preference. For example, a question such as *how long could this possibly take for a worker in your position to complet*e may be more effective to support realistic daily planning.

4.8 Discussion of Study 4

This study examined the extent to which current personal task management (PTM) applications support evidence-based strategies for improving time estimation accuracy. Building on the four strategies identified in Study 3: duration feedback, distributional data, task breakdown, and induced neutrality, this study provides a systematic analysis of 47 popular PTM tools to evaluate whether and how these strategies are supported in practice. The results highlighted a significant disparity between psychological insights into time management and their practical implementation in productivity tools. Despite the high prevalence of subtasks across apps, there were limitations in their alignment with the strategy of task breakdown, as recommended by psychological literature. Moreover, the implementation of the feedback strategy was found to be limited, with very few applications offering a suitable fit with time tracking and analytics functionalities that were easy to use. Efforts to integrate distributional data and induced neutrality within existing apps were unsuccessful.

Inaccurate time estimations pose a significant challenge in time management, yet this aspect remains largely overlooked in traditional time management literature. This study introduces a novel perspective by emphasizing the importance of correcting time estimations, akin to enhancing realistic planning. While HCI research supports the efficacy of feedback and task breakdown [78], these concepts are not prominently featured in the literature guiding app development as *bias reducing*, which often prioritizes facilitating *execution*. The apparent lack of awareness among users and designers regarding the value of debiasing for improving planning skills may explain the underutilization of such functionalities. By advocating for the enhancement of realistic planning through improved time estimations, apps could potentially offer more tailored analytics on time usage and incorporate user-friendly features like proactive and semi-automatic time tracking for individuals within PTM apps.

The addition of more features to apps can potentially overcomplicate design. However, the study does not advocate for these features to be universally present at all times. Rather, they serve as "coaching" features that individuals may find beneficial during periods of habit change or heightened stress. While the study primarily assesses the presence and ease of use of functionalities aligned with specific strategies, future research could delve into enhancing the user-friendliness of these features and integrating them seamlessly into existing apps. The rest of the studies in this thesis provide insights into the effects of debiasing in practice. With this work, I aspire to inspire further exploration and innovation in this area. The lack of robust time estimation strategies in current PTM apps emphasizes the pressing need for a paradigm shift in their design to better support users in addressing time estimation biases and improving their productivity in a way that is useful for them.

This paper additionally contributes methodologically by advancing the approach to functionality review studies. Initially, I conduct a comprehensive analysis of all functionalities before narrowing down to those of interest, thereby reducing selection bias. The approach involves downloading and reviewing apps firsthand rather than solely relying on apps descriptions, a method seldom utilized in prior studies. Descriptions often prove inaccurate due to apps making exaggerated claims, hence direct evaluation ensures a more precise assessment of capabilities. Moreover, evaluating apps firsthand allows for additional observations, aligning closely functionalities with strategies.

Several measures were taken to ensure objectivity, including a second-rater reliability assessment for the corpus and ongoing quality assurance procedures through research group discussions. Despite these efforts to ensure objectivity, there still may be limitations. For example, the final evaluation step was conducted by a single researcher. Additionally, while we updated the evaluation at the end of 2023, it's important to acknowledge that apps undergo frequent updates and changes, meaning our assessment provides only a snapshot to guide future efforts.

The advocacy for realistic planning strategies in this study may raise concerns about the potential impact on users' motivation. Some may argue that overly realistic planning could lead to reduced motivation to complete tasks promptly. However, we contend that striving for accuracy in task estimation is desirable, as it allows for effective prioritization and the maintenance of challenging yet achievable goals [25]. By ensuring that estimates remain within a challenging range, users can better manage their workload and mitigate the negative consequences of procrastination and delays. The rest of the thesis is focused on understanding the effects of increasing realism of plans in the real-world.

4.9 Summary of Chapter 4

This chapter presented Studies 3 and 4, which together examined how evidence-based debiasing strategies can be translated into design knowledge for planning technologies. Study 3 synthesised findings from 28 experimental studies in psychology and HCI, identifying four strategies for improving time estimation accuracy: duration feedback, distributional data, task breakdown, and induced neutrality. Building on this framework, Study 4 conducted a functionality review of 47 personal task management (PTM) applications to evaluate whether these strategies are supported in practice. The analysis showed that while some features, such as time tracking, Pomodoro timers, and analytics, could provide duration feedback, and subtasks or templates could support task breakdown, there was little support for distributional data and no evidence of features addressing induced neutrality. Even where relevant features existed, they often placed a high burden on users and were not designed with debiasing in mind.

Together, the two studies reveal a substantial gap between psychological evidence and the design of PTM tools. Beyond focusing on task execution, planning technologies could also incorporate science-informed features that actively support realistic planning. These insights provide concrete design directions and establish the foundation for the field interventions reported in the next two chapters.

Chapter 5

Studies 5 and 6: Debiasing in the field

Parts of this chapter have been published in Ahmetoglu, Brumby and Cox [140].

Chapter Outline *This chapter reports two field interventions testing task duration feedback in academic work: Study 5 used a commercial app and Study 6 a spreadsheet tracker. Together they show how feedback influences planning accuracy, productivity, and daily experiences, while also revealing its limits as a debiasing strategy.*

This chapter addresses the third sub-research question in this thesis focused on understanding debiasing in practice, moving from identifying evidence-based strategies and evaluating existing tools to testing interventions in real-world contexts. Study 5 deployed a two-week intervention with 10 participants using Sunsama, a commercial personal task management (PTM) app. Quantitative analyses showed no measurable improvements in planning accuracy. Qualitative findings revealed mixed experiences: while some participants reported greater awareness of time use, reduced over-optimism, and more motivation to plan, most did not consistently engage with Sunsama's time-tracking and ritual features. Study 6 built on these findings by addressing limitations of the app-based approach through a four-week spreadsheet-based intervention with 30 participants. This manual method supported more consistent engagement with feedback and allowed users to directly compare planned and actual durations. While survey results again showed no significant improvements in objective accuracy, participants reported increased self-assessed realism of task duration estimates, decreased procrastination, higher self-efficacy, and improved perceived control over time. Interview data revealed that feedback helped participants recalibrate expectations and set more realistic goals.

5.1 Introduction to Study 5

This chapter presents the first of two field intervention studies designed to investigate how digital tools can support more accurate planning in academic knowledge work. In this study, I take the next step toward answering this question by testing one of the most promising strategies identified in previous chapters: providing feedback on the duration of tasks.

Earlier chapters established the significance and complexity of task planning in academic work. Studies 1 and 2 showed that inaccurate time estimation is widespread and contributes to disengagement from planning. These studies addressed sub-question 1 by identifying the cognitive and contextual barriers academic knowledge workers face, such as unrealistic expectations, vagueness in daily plans, and lack of reflection. Study 3 responded to sub-question 2 by synthesising experimental evidence on methods to improve time estimation accuracy and identifying four debiasing strategies: duration feedback, distributional forecasting, task breakdown, and induced neutrality. Study 4 then evaluated how well these strategies are supported in existing personal task management (PTM) tools, revealing a major design gap, particularly in the provision of feedback features. These findings highlight the need for interventions that can be practically implemented and integrated into everyday planning routines.

This chapter addresses sub-question 3: *What are the effects of task planning debiasing interventions on planning and productivity outcomes?* Specifically, I evaluate the effects of task duration feedback in a real-world setting. This strategy was selected for field testing because it was the most commonly supported feature among the tools reviewed in Study 4 and has strong backing in the psychology literature for reducing bias in experimental settings. However, whether such feedback is effective in everyday academic contexts, and how individuals engage with it over time, remains poorly understood.

To explore this, the study involved a two-week field trial using Sunsama, a feature-rich PTM tool selected for three reasons. First, unlike many other apps, Sunsama includes time tracking and daily review rituals that allow users to reflect on how long tasks actually took compared to their estimates. This makes it a rare commercially available tool that partially implements the duration feedback strategy identified in Study 3. Second, it offers structured planning workflows (including a daily planning ritual and review prompts), which aligns with the finding from Studies 1 and 2 that reflective structure can support re-engagement with planning. Third, Sunsama was selected for ecological validity: it is a real tool used by professionals and integrates with other productivity systems, providing a realistic testbed for evaluating planning support in everyday academic work.

5.2 Method of Study 5

To investigate the research questions, namely, explore how task duration feedback can reduce the optimistic planning bias and when, and in what ways such feedback is beneficial, this study employs a commercially available to-do list app. The app includes time tracking and feedback features, which have been hypothesized to mitigate planning bias in academic work [100, 77].

5.2.1 Participants and recruitment

Ten taught postgraduate students from a UK university took part in a 2-week long field intervention study. Their mean age was 24 years, with $N = 8$ identifying as female and $N = 2$ identifying as male. They were recruited through the graduate program university network. The study was advertised as an opportunity to improve time management and did not require any prior experience to join. They were compensated with 20 pounds for their time. The study was approved by UCL Ethics Committee.

5.2.2 Intervention design and procedure

In April 2024, participants attended one of three small-group (2 - 4 people) online workshops. Each two-hour session included an initial survey, a discussion on planning habits and productivity challenges, and an introduction to Sunsama's time-tracking and ritual features. To minimize stress, participants were encouraged to focus on self-awareness and well-being rather than strict time management.

Sunsama is a digital planner designed to help users set and track daily plans. Participants engaged with three key features: daily planning, daily review (collectively referred to as rituals), and a task-tracking timer. The daily planning ritual involved reviewing unfinished tasks from the previous day, adding new tasks, assessing workload, and prioritizing tasks. Time tracking allowed users to log time spent on tasks using a start-stop timer or manual entry. The daily review ritual provided a structured reflection process, summarizing completed tasks and displaying a breakdown of time spent across different activities. Figures 1(a)(c) illustrate these features.

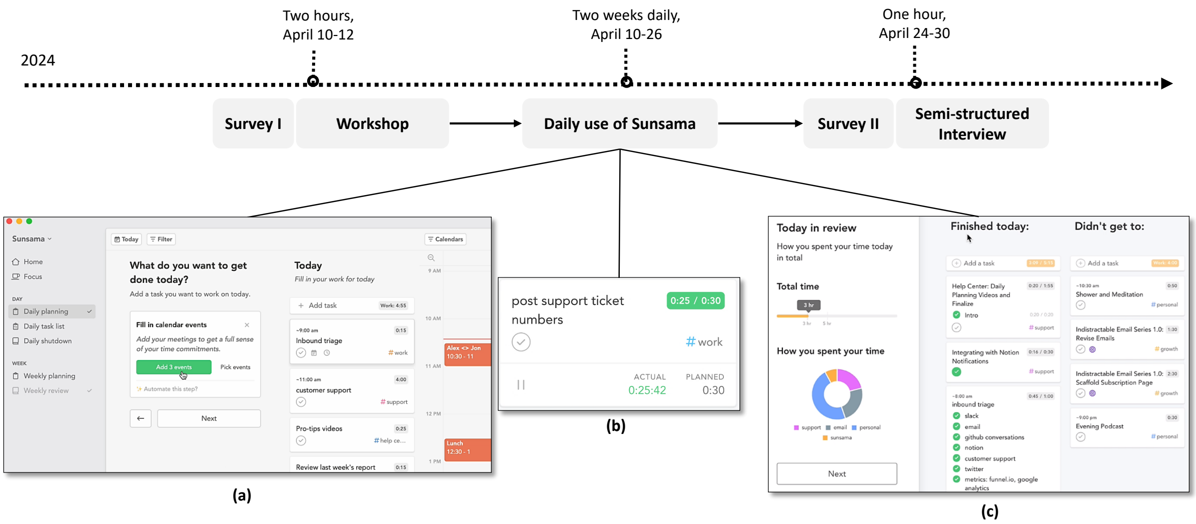


Figure 5.1: Study 1 procedure: Participants first attended an introductory workshop and completed Survey I. They then used Sunsama for two weeks, with encouragement to engage in (a) the daily planning ritual, (b) time tracking, and (c) the daily review ritual. Finally, participants attended end-of-study interviews and completed Survey II.

While participants were introduced to these features during the workshop, their use could not be enforced. Sunsama does not require time tracking, but regardless of whether participants logged time, they were shown the rituals features upon launching the app each day. This included reviewing completed and incomplete tasks from the previous day and planning for the next (Figure 1c). If time tracking was used, this data was displayed alongside a pie chart of time distribution. Users could bypass the rituals screen and go directly to their daily task list. To maintain ecological validity, participants were free to engage with these features as they naturally would in real-world use.

After the workshop, participants engaged with Sunsama for two weeks. At the end of this period, they took part in interviews where they also filled out the survey for a second time. Figure 1 illustrates the procedure described above.

5.2.3 Outcome measures: optimistic bias, time management outcomes and subjective experiences

To measure change in time estimation skills, the survey included task duration estimation questions, such as time to read a research article or complete coursework. Time estimation tasks were modeled after previous studies on the optimistic planning bias [60], with the expected outcome being a longer predicted duration post-intervention. The survey also incorporated broader time management measures, previously linked to such outcomes in research, to explore the wider impact of debiasing beyond estimation accuracy [4, 74, 75]. Trait procrastination was measured with a 5-item scale [141] with example items such as *I frequently find myself putting important deadlines off*. General self-efficacy was measured with a 6-item scale [142] with example items such as *No matter what comes my way, I'm usually able to handle it*. Perceived control over time was measured with a 4-item scale [94] with example items such as *I feel in control of my time*. Responses were recorded on a 5-point Likert scale ranging from Strongly Disagree to Strongly Agree.

Semi-structured interviews were conducted to complement the survey data. The interview script had questions about the features that were used over the last two weeks (e.g. which ones, how frequently throughout the two-week period they were used, the level of detail of tasks), and the impact the app had on productivity and well-being.

5.2.4 Data analysis

Survey and time estimation task data met the assumption of normality and parametric tests were applied. The data were analysed through paired-sample *t*-tests. The values from the Likert scale were summed for each participant. The interview data was analysed with inductive thematic analysis, following the recent guidelines of Braun and Clarke [143, 144]. One of the authors repeatedly reviewed interview transcripts to familiarize themselves with the data and made initial notes on recurring ideas. Using NVivo, segments of text relevant to the research questions were flexibly coded, reflecting participants' language. Codes were iteratively refined and grouped into distinct themes to capture significant patterns. Research group discussions were held to present findings and gather feedback, ensuring the themes were coherent and well-developed.

5.3 Results of Study 5

5.3.1 No differences in time estimations and time management outcomes

One participant withdrew due to an unrelated health emergency, leaving data from nine participants for analysis. Average time estimates for different academic tasks were compared before and after the intervention. For reading a research article, participants initially estimated 50 minutes (SD = 31.62) in Survey I and 61.25 minutes (SD = 50.67) in Survey II, with no significant difference between these estimates ($t(8)$ = -0.869, p = 0.414). Data from four other estimation tasks was excluded from analysis as participants provided vague responses (e.g., *a few days*) rather than concrete time predictions.

Average trait procrastination values were 17.37 (SD = 5.29) before and 15.25 (SD = 4.84) after the intervention, average perceived control of time values were 16.5 before (SD = 2.92) and 18.25 after (SD = 3.15), and average general self-efficacy scores were 19.62 before (SD = 3.71) and 20.37 after (SD = 3.08). There were no statistically significant changes in trait procrastination ($t(8) = 6.5$, $p = .201$), perceived control of time ($t(8) = 2$, $p = .072$), and general self-efficacy ($t(8) = 10.5$, $p = .313$) scales.

5.3.2 The experience of using Sunsama

Theme 1: Low Engagement With Time Tracking and Rituals.

Most participants did not engage with the recommended functionalities which they were instructed to use during the workshops. The majority of participants (6 out of 9) did not consistently use the time tracking timer or rituals, and shared that they preferred to **use the app as a to-do list**. As P3 noted: *I didn't use the time tracking - I used it as a to-do list. If I hadn't done something then I would drag it to the next day.*

One common barrier to engaging with time tracking was that after starting to track a task, participants often **forgot to pause the timer** if they took a break or to stop it if they switched tasks: *I also tried the timer and the Pomodoro timer once but and I lost lost track of that. It ended up counting for the entire day* (P5).

Participants shared that the rituals notification came at an **inconvenient time** and they decided to close the pop-up. P1 shared that: *I don't think it was helpful for me because I just closed it when it popped up because I was busy with work and I did not want to be interrupted.* The review ritual was often disregarded because participants stayed up late past midnight and were not ready to wrap up their work. *I kept to the basics because the [review] wasn't the time I shut down at because I do work until late so I ignored the [review]* (P2). When participants were busy with coursework, they frequently disengaged from planning altogether: *I did not look at it as much during the second week when I had two coursework deadlines* (P7).

Participants experienced a **steep learning curve** with the app. *Even though the app is quite neat and minimalistic, it is still a new app to add to your routine and I think there was a barrier of getting used to it* (P8). Participants frequently reported spending the first week exploring and getting used to the app. At times, they also did not recall the instructions of the study which left them unsure about which features to use: *Seeing all the different features, I honestly could not remember how I was supposed to use it, so I did what I normally do [...] I made lists*. (P9). Remembering to use the app everyday was a challenge. P5 noted: *Honestly, I couldn't remember to do it most of the days. I'd remember halfway through the day, but then the next day I'd forget to do it in the morning*.

Some of the participants, who did use time tracking or rituals more consistently, reported finding these features *very helpful for* ***increasing awareness***. P1 reported improving her understanding of how much could actually be achieved in a day: *It makes me feel like I know my capabilities more realistically* (P1). The rituals feature was also found helpful to assess progress towards goals and avoid overloading the day with too many tasks. *It gave me a very clear visual to see how much time I had left of my day, it made it easier for me to estimate, OK, today I can do approximately 3 tasks* (P4). Tracking helped participants to realize that aiming for less resulted in more efficient work: *It helped me realize that it's easy to get stuff done when there's less to do* (P6). The data visualisations were found useful, too: *I found super helpful to see how I aligned with my goals with the pie-chart.* (P4)

Theme 2: Mixed Success of Coping With Overloaded To-Do Lists.

Participants used Sunsama mainly as a to-do list tool and they noticed that they were not completing the tasks they were setting out to complete each day. As a result, they reported increased self-awareness toward their tendency to overload their to-do lists, and this realization had a mixed impact.

For some, seeing the tasks that were not achieved was disappointing and made them less engaged in planning. *It is a bit discouraging when I set myself that I'm going to do this and it ends up not being realistic, so I don't end up doing it* (P3) and demotivating: *I think there are times where I wasn't successful in completing my tasks, so I was less motivated* (P4). These feelings were especially unpleasant during stressful periods when deadlines were approaching, and the app served as a **negative reminder of not being on track**. As a result, participants reported **disengaging from planning** altogether: *When I'm super stressed, I don't really want to be planning. I just want to get into work because I may not get to finish the task and that makes me feel bad about myself* (P2). Some shared that the failure to meet plans meant that planning was not working for them: *I found that the actual tasks did not match what I was predicting at all, not only in terms of time, but also in terms of what am I actually doing that day so, the system was very quickly falling for me.* (P7)

Some participants reported overcoming the initial disappointment by realizing that there are other **external factors** at play that lead to overloaded to-do lists apart from not being sufficiently productive. P8, for example, who had a full-time job alongside her graduate degree, realized that her workload was too high to be able to manage everything on time: *I think because I was tracking the time, I realized how much I really did have to get in a day and the app did tell me that it looks a little bit unreasonable to do [...] you realize maybe this is not a me issue, it's more the workload issue*. Others realized that they were trying to be *perfect* when, in fact, the plans were flawed to start with, and were leading to disappointment. P4 illustrated her experience: *I realized that I feel disappointed in myself. And then I realized that this isn't really realistic and that it's not something to take personally.*

Other participants reported **starting new strategies** to make the experience more *manageable* and *to feel better about [themselves]*. Breaking down tasks into smaller steps was one of these strategies: *I realized I couldn't finish all of my tasks because they were too big. I started setting smaller tasks so I could tick them off and that made me feel good* (P2). Others started to use prioritization: *The app forced me to think about the essential things even if just in my head* (P5). Some reported positive impacts, such as feelings of more control over time and having structure as a result: *Daily planning was really helpful, I felt more in control and it gave a nice structure to my day* (P5). Using the app also encouraged them to work more efficiently: *Of course, I probably worked a bit harder, but not because of the study, it was more to show myself that I can meet my plans* (P8).

However, the efforts to handle overloaded to-do lists were **not always successful.** Even though some tried to prioritize and break down tasks, they still found themselves unable to meet what they thought would be more realistic targets: *I was failing at doing certain tasks, it made me more realistic and realize that like I can't complete 11 tasks today. And then, I learned to prioritize [...] And then feeling disappointed because I can't do one of the things* (P4).

5.4 Discussion of Study 5

The results of Study 5 indicate that Sunsama's task duration feedback had no significant effect on improving participants' accuracy in predicting task duration or on broader time management outcomes. Although participants' time estimates showed a slight upward trend, this change was not statistically significant. Interview data revealed that most participants did not consistently engage with the app's time-tracking features, instead using Sunsama primarily as a to-do list. This limited engagement likely explains the absence of significant effects in the self-report measures.

These findings help explain why many planning technologies face high dropout rates [54]. Several participants expressed frustration or discouragement when their optimistic plans failed to unfold as intended. This reinforces insights from Studies 1 and 2, which identified the emotional impact of unmet expectations as a key barrier to sustained planning engagement [78, 77]. This study extends previous work by highlighting a fundamental gap in many PTM tools: the implicit assumption that users already possess accurate time estimation skills. In contrast, our findings suggest that without appropriate scaffolding, users may abandon tools that surface inaccuracies without providing the necessary support to manage their effects.

Among the minority of participants (3 out of 9) who did engage with the time-tracking features, not just the task list, most reported positive outcomes. These participants described setting more realistic daily goals after receiving feedback on how long tasks actually took. For example, one participant found that planning for less work each day led to a greater sense of accomplishment and, paradoxically, to completing more. This suggests that task duration feedback can help recalibrate expectations and foster healthier planning habits. Importantly, participants viewed this recalibration as a relief rather than a failure, an opportunity to shift toward more sustainable productivity. However, the overall lack of engagement highlights that, in real-world conditions where app use is voluntary, most users are unlikely to naturally adopt feedback-driven planning tools. As such, these findings point to the importance of exploring alternative methods for delivering feedback that better align with user habits and motivations.

5.4.1 Implications for research

The first implication from this study is that feedback-based planning interventions must begin by addressing the negative emotional responses that users often experience when plans go unmet. One way to do this is by avoiding the requirement to plan upfront. Instead, interventions could focus solely on feedback provision, with optional functionalities, such as task prioritisation, reflection, or breakdown, offered as additional support rather than enforced steps. Participants may also benefit from onboarding materials that explicitly address the emotional experience of planning and offer coping strategies.

Second, to minimise planning disengagement, interventions should avoid overwhelming users with complex or unfamiliar interfaces. Designs that are simple, minimalist, and familiar, such as spreadsheet- or document-style layouts, may help reduce cognitive friction and lower the barrier to entry [117]. Third, future tools could do more to embed feedback into users' existing workflows, increasing the likelihood of habit formation. This might include reminders or contextual prompts that support consistent time tracking and reflection as part of the daily routine [126].

Finally, a limitation of this study was the reliance on subjective time estimation, which often produced vague or approximate responses. Many participants did not provide specific time predictions, making it difficult to assess changes in accuracy over time. Future studies could address this by using more structured estimation tasks and by comparing predictions with actual task durations.

5.5 Summary of Study 5 and bridge to Study 6

The findings from Study 5 make three contributions to the thesis. First, they provide empirical evidence on the use of task duration feedback in everyday planning. While objective improvements in estimation accuracy were limited, participants who regularly engaged with the feedback reported increased awareness of time use, reduced over-optimism, and stronger motivation to plan. Second, the study highlights contextual and design barriers that constrained engagement: many participants did not use the features consistently due to workflow misalignment or perceived effort. These results underline the need for feedback mechanisms that integrate more naturally into daily routines and minimise user burden. Third, the study sets the stage for Study 6, which responds to these limitations by introducing a simplified, manual time-tracking intervention designed to sustain engagement. Together, these findings contribute to the broader aim of the thesis: to inform the design of planning technologies that are both psychologically grounded and practically workable in knowledge-intensive settings.

5.6 Introduction to Study 6

Here, I present the second of two field intervention studies aimed at exploring how task duration feedback can reduce the optimistic planning bias in academic knowledge work. Together with Study 5, this study addresses the third sub-question of the thesis: *What are the effects of task planning debiasing interventions on planning and productivity outcomes?* The first field study (Study 5) provided initial evidence that duration feedback has the potential to support more realistic planning by helping participants adjust their expectations and reduce perceived over-optimism. However, it also revealed significant barriers to engagement: most participants used the PTM app (Sunsama) primarily as a to-do list and did not regularly interact with its time-tracking and reflective features. These limitations restricted the impact of the intervention and highlighted the importance of designing feedback mechanisms that are simple, unobtrusive, and better integrated into users' routines.

To build on these findings, this follow-up study introduces a more lightweight, flexible intervention using manual time tracking in a spreadsheet format. This method was chosen to reduce friction and increase user control: factors identified in the previous study as important for engagement. By allowing participants to enter and compare planned versus actual task durations in a format familiar to most knowledge workers, the study aims to provide duration feedback without relying on learning how to use a feature-rich app. Additionally, this study expands the sample size and duration of the intervention to increase generalisability and capture more meaningful shifts in behaviour over time. Specifically, Study 6 explores the main research question for this part of the thesis by looking at 1) how task duration feedback can reduce optimistic planning bias, 2) when it is perceived as beneficial, and 3) in what ways it supports productivity and everyday planning experiences. These questions are revisited in the Discussion section of Study 6.

A key improvement in this study is the inclusion of both objective and perceived measures of optimistic planning bias. Whereas Study 5 relied primarily on self-report data and informal comparisons, this study collects structured estimates and actual durations for a set of tasks, allowing for a clearer assessment of accuracy. It also incorporates again the same validated measures of broader time management outcomes, such as procrastination, self-efficacy, and perceived control over time, to examine whether changes in planning behaviour are accompanied by changes in productivity and wellbeing. Finally, semi-structured interviews provide qualitative insight into participants' experiences with the intervention, including how they interpreted and responded to the feedback, and what barriers or enablers they encountered. This mixed-methods approach enables a deeper understanding of the mechanisms through which feedback affects planning, as well as the contextual factors that shape its impact.

5.7 Method of Study 6

5.7.1 Participants and recruitment

Thirty participants took part in this study. They were PhD students at a UK university. Their mean age was 33 years, with N = 23 identifying as female and N = 7 identifying as male. Twelve were in their first year, four in their second year, six in their third year, and eight in their fourth year. Participants represented a wide range of fields, including psychology and social sciences, economics, biology, neuroscience, education, chemistry, human-computer interaction, physics, medical sciences, and history. Participants were invited to participate via emails sent to university mailing lists and departmental communication channels. The study was advertised as an opportunity to improve time management and did not require any prior experience to join. Participation was rewarded with 55 pounds. The study was approved by the UCL Ethics Committee.

5.7.2 Intervention design

Following insights from the first study, this intervention aimed to *minimize negative emotional responses* to unmet expectations by providing *feedback only* rather than requiring participants to make plans. A manual time-tracking approach was chosen to avoid the high burden of timer-based tracking while allowing *optional customization* (e.g., prioritization, reflection, task breakdowns) through simple edits in a spreadsheet. The tracker, implemented in Microsoft Excel and trialed by a researcher for four weeks, was adaptable to participants' preferences. They could also record reflections freely, though this was not required to prevent negative reactions.

To provide *explicit training in coping with self-tracking*, participants created example tracking days during onboarding workshops. They practiced three scenarios: a standard day (*usual level*), a day with minimal tracking during overwhelming periods (*low level*), and a day with detailed tracking of breaks and specific tasks (*high level*). This approach, informed by research on implementation intentions [31] and coping planning [145], helped participants prepare for varying work conditions. They were encouraged to track at least at the *usual* (or optionally *high*) level for two of the four weeks and maintain at least *low* tracking for the rest, with the option to contact researchers if needed.

To minimize confusion, we opted for a *simple, familiar, and flexible tool* with minimal learning curve. A pre-filled example day was included, and participants could transfer logs from other tools

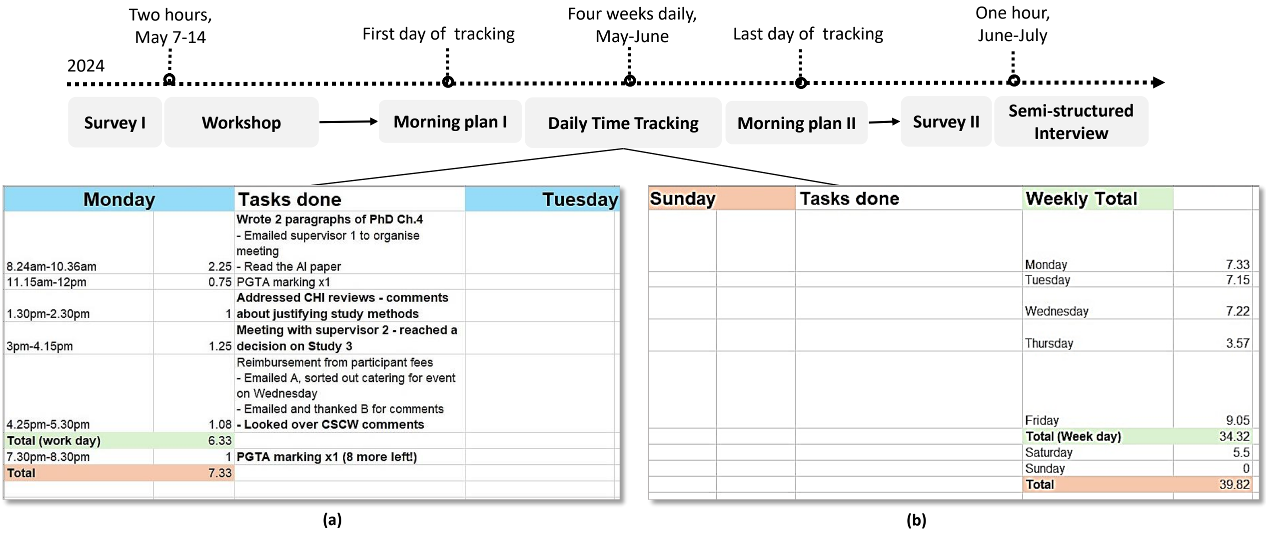


Figure 5.2: Study 2 procedure: Participants first attended an introductory workshop and completed Survey I. They then used a spreadsheet time tracker for four weeks, manually tracking time spent on tasks: (a) shows an example day of tracking included in the spreadsheet for guidance and (b) example weekly total hours. On the first and last day of tracking, they filled in a morning plan sent back to the researchers. Finally, participants attended end-of-study interviews and completed Survey II.

(e.g., paper, digital trackers) at their convenience. The spreadsheet featured daily columns for task logs, start and end times, and a "Tasks Done" section. A "Work Day Total" calculated hours worked, with weekly summaries for workdays, weekends, and overall totals (see Figure 5.2).

To *support habit formation and seamless feedback*, participants logged time at *natural break points*, such as before and after work blocks, rather than after each task switch. This approach, informed by research on minimizing interruptions [146], aimed to integrate tracking into existing routines. For *reminders*, participants selected their preferred notification method and frequency (e.g., Teams message, email, calendar event), and researchers sent reminders accordingly. To reduce demand characteristics and encourage authentic tracking [147], participants were not required to share their spreadsheets. At the end of the study, they had the option to share if they felt comfortable.

5.7.3 Procedure

The study procedure is illustrated in Figure 5.2. Participants first attended one of nine small-group (36 participants) two-hour workshops, which introduced the study, included an initial survey, discussions on time management habits, and a demonstration with hands-on practice of the tracking method. They then used the self-tracking spreadsheet daily for four weeks, choosing their own start date after the workshop. On the first and last day of tracking, they completed morning plans estimating task durations. At the end of the study, they completed a second survey and participated in interviews, where they reviewed their two morning plans and filled in actual task durations based on their spreadsheet logs, and classified each task as optimistic, realistic or pessimistic with the help of the researchers.

5.7.4 Outcome measures: optimistic bias, time management outcomes and interviews

Optimistic bias was measured using two complementary indicators: *discrepancies* between estimated and actual task durations recorded in participants' spreadsheets and *perceived bias* based on their own evaluations during interviews. The first indicator, adapted from the augmented diary method [15, 77], provided an objective comparison of estimated versus actual task durations and total workday length, identifying patterns of over- or under-estimation. However, this quantitative measure lacked contextual nuance. To address this, the second indicator involved participants reviewing their own data and classifying each estimate as pessimistic, realistic, or optimistic. This approach accounted for contextual factors, such as whether a shorter-than-expected task duration reflected a pessimistic estimate or an unfinished task, which would indicate an optimistic estimate. While subjective, the perceived bias measure was grounded in the discrepancy data, offering a more context-aware complement to the first indicator.

The measure of broader work outcomes was the same as in the first study, focusing on trait procrastination [141], general self-efficacy [142], and perceived control over time [94]. Participants also took part in semi-structured interviews lasting 60 - 90 minutes to explore their experiences with time tracking. These interviews covered engagement with tracking (e.g., methods, evolution, and reflections), its effects on time management and planning behaviors, and any changes in perceptions of task durations. Participants also shared suggestions for improving the design and usability of time-tracking tools.

Data Analysis

Morning plan data was processed by excluding personal tasks (e.g., showering), canceled tasks, and those without clear time records. Survey and time estimation data met normality assumptions, allowing for parametric tests. Estimated and actual task durations were totaled for each participant

(a)

Task	Estimated duration	Actual duration	Optimistic, realistic or pessimistic
Gather documents for [anonymous] and PhD work	30m	30m	Realistic
PhD work for [anonymous project]	60m	75m	Realistic
Agenda for meeting with [anonymous]	30m	60m	Optimistic, because it took longer to complete on that day
Email file for review	30min	0m	Optimistic, it was not done on that day
Unplanned tasks	-	-	-
Total without unplanned	150m	165m	-
Total with unplanned	150m	165m	

(b)

Task	Estimated duration	Actual duration	Optimistic, realistic or pessimistic
Data into Nvivo: First upload	60m	60m	Realistic
Plan second upload Second upload	90m	130m	Realistic, because there were interruptions
Address paper comments	90m	30m	Optimistic, because it took several days to complete
Unplanned tasks	-	-	-
Team meeting	-	90m	-
Total without unplanned	240m	220	-
Total with unplanned	240m	310m	-

Figure 5.3: Tables used during the interviews with a participant that measured optimistic planning bias based on data from participants' (a) first and (b) second morning plans with inputted actual durations from their trackers.

on the first and last day of tracking to calculate objective bias and compare changes. The number of optimistic, realistic, and pessimistic estimates per participant was also analyzed, comparing overall category shifts before and after the intervention, as well as direct transitions between categories. Figure 5.3 illustrates an example from the dataset used for analysis. To control for Type I errors, a Bonferroni correction was applied when multiple tests were performed on the same dataset. Survey and interview data were analyzed using the same approach as Study 1.

5.8 Results of Study 6

5.8.1 Engagement with the study

Valid Responses.

Of the 30 participants who attended the workshops, two dropped out: one did not respond after the tracking period, and the other withdrew due to illness. Of the remaining 28, three provided insufficient tracking details to measure actual task durations from their morning plans. As a result, interview data from 28 participants was analyzed, while the quantitative analysis included data from 25 participants.

Engagement with time tracking

Interview data indicated that most participants (N = 19) tracked their time for all four weeks, while three continued for five weeks. Six participants tracked for three weeks due to a planned week off or

(a)

Wednesday		Tasks done	Thursday		Tasks done
07.57- 08.31	34	Admin, e-mail, Twitter	08.29 - 08.58	29	Admin, e-mail, Twitter
08.32 - 09.49	77	[anonymous] slides	09.02 - 09.43	39	[anonymous] analysis
10.17-12.40	97	[anonymous] analysis	10.05-11.24	79	[anonymous] analysis
13.18-13.27	9	Admin, e-mail etc.	11.30-11.47	17	email & seminar arrangements adr
14.04-15.04	60	[anonymous] analysis	12.12-12.58	46	[anonymous] prepare
15.16-15.53	37	[anonymous]slides prepare	13.00-13.21	21	[anonymous] analysis
16.29 - 16.42	13	Admin	14.45-15.02	17	[anonymous]conference planning
16.43 - 17.53	70	[anonymous] meeting prepare	15.08-15.53	45	[anonymous] analysis
20.21 - 2:32 + 20:44 - 20:59 + 21.07 - 21.46.	71	Student slides	16.12 - 16.59	47	[anonymous] flights sort
			17.00 - 18.10	70	[anonymous] meeting
Total (work day)	7.80				
GOALS	y	1 x systematic review	Total (work day)	6.83	
	y	1 x check other systematic reviews			
	n	[anonymous] investigate	GOALS	y	1 x systematic review
	y	Meeting with [anonymous]		y	1 x check back reviews 1-6
	y	prepare student MsCslides		n	submit review doc to [anonymous]
					[anonymous]
				y	Private [anonymous]meeting

(b)

Thursday		Tasks done	Friday		Tasks done
10:00 - 10:45	0.75	podcast meeting	10:30-1:30	3	seminar; emails etc
11:00 - 2:00	3	induction + emails + meeting	2-2:30	0.5	meeting with [anonymous]
2:00 - 2:30	0.5	meeting with [anonymous]	2:30-5:30	3	literature
2:30 - 5	2.5	organizing literature, R code files			
total	6.75		total	6.5	

(c)

Weekly Total	
Monday	8.5
Tuesday	7
Wednesday	7.75
Thursday	4.5
Friday	1.25
Total (Week day)	29
Saturday	0
Sunday	0
Total	29

Figure 5.4: Self-tracking data gathered by participants. (a) High level of detail self-tracking data with customization of colours and goals. (b) Lower level of detail self-tracking data with no customization to the spreadsheet. (c) Weekly total hours tracked

non-academic commitments. Twenty participants began tracking on a Monday, with the rest starting on different weekdays (Tuesday: $N = 1$, Wednesday: $N = 2$, Thursday: $N = 4$, Friday: $N = 1$).

Participants employed various tracking methods, differing in detail and frequency. Some logged each task with precision, recording durations from 5 minutes to several hours. Others grouped tasks into broader sessions, typically 15-30 minutes to 2-3 hours, often marked by breaks. Some used fixed intervals of 30-60 minutes, either noting tasks on a calendar and transferring them later or recalling and logging them at the end of the day. In some cases, tasks were initially scheduled in approximate time blocks, with adjustments made afterward to reflect actual time spent.

Some participants customized the spreadsheet, often by color-coding tasks, highlighting key activities, or adding to-do lists. For instance, some marked delayed tasks in red or used different colors to distinguish between PhD work and paid work. Others added to-do lists below daily logs, starting each day with a plan and later comparing it to their tracked work. Figure 5.4 presents two examples from participants' spreadsheets (with permission): (a) a detailed version with customizations, and (b) a higher-level summary, both resulting in similar weekly hour totals shown in (c).

Eight out of 28 participants used additional tools alongside the spreadsheet, including time-tracking apps like Toggl (https://toggl.com), calendars, or paper diaries. These tools were chosen for familiarity and convenience. Participants primarily logged time using these methods and later transferred the data to the Excel spreadsheet, typically at the end of each day or week. Most participants preferred weekly reminders ($N = 15$), with email ($N = 7$), calendar invites ($N = 5$), and Teams messages ($N = 3$) as their chosen methods. Daily reminders were selected by five participants (four via email, one via a calendar invite), while eight participants opted for no reminders.

5.8.2 Changes in the optimistic bias

Discrepancies in Estimated and Actual Durations

Eleven tasks were excluded from the first diaries (six canceled, three personal, and two without time records), and fifteen from the second diaries (six canceled, five personal, and three without time records). This resulted in 150 task duration estimates (75 before and 75 after the intervention). On average, participants provided 3.64 estimates in the first diary and 3.60 in the second.

In the first diary, participants planned to do tasks with an average total estimated duration of 5 h 55 min ($SD = 131$ min) and reported spending 5 h 34 min ($SD = 117$ min) working on those tasks,

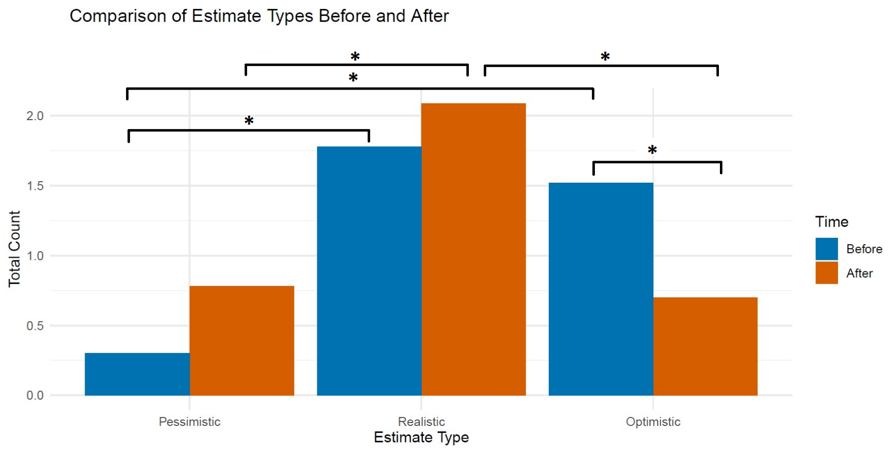


Figure 5.5: A bar chart showing the distribution of optimistic, realistic, and pessimistic estimates before and after the intervention.

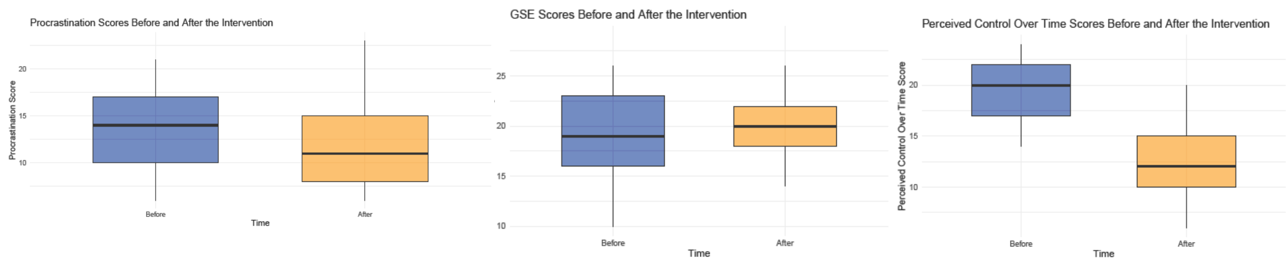


Figure 5.6: Box plots showing comparison of scores before and after the intervention for procrastination (left), general self-efficacy (center), and perceived control over time (right). Boxplots display the median (horizontal line), the interquartile range (box), and whiskers extending to the smallest and largest values within 1.5ŒIQR.

and the difference was not significantly different (t = 0.72, p > .05). They also spent 20 min (*SD* = 37 min) on unplanned work tasks. Together with unplanned work, the total workday duration amounted to 6 h 2 min (*SD* = 131 min). In the second diary, participants planned to do tasks with an average total estimated duration of 5 h 19 min (*SD* = 123 min) and reported working on these tasks for an average total duration of 4 h 32 min (*SD* = 131 min) which was significantly different indicating bias (t = 3.16, p = .004). They also spent 48 min on unplanned work tasks (*SD* = 60 min). Together with unplanned work, the total workday duration amounted to of 5 h 6 min (*SD* = 130 min) in the second diary.

Perceived reduction in optimistic bias

Before the intervention, participants made an average of 0.30 pessimistic estimates (SD = 0.45), 1.78 realistic estimates (SD = 2.09), and 1.52 optimistic estimates (SD = 0.70). A repeated-measures ANOVA found significant differences among these categories ($F(2, 48) = 17.12$, $p < .001$). Post-hoc pairwise comparisons showed that pessimistic estimates were significantly lower than both realistic ($p < .001$) and optimistic estimates ($p < .001$), while no significant difference was found between realistic and optimistic estimates ($p > .005$). This suggests that participants tended to make either realistic or optimistic estimates rather than pessimistic ones, indicating a bias toward optimistic planning.

After the intervention, the average number of pessimistic estimates increased to 0.78 (SD = 0.86), realistic estimates to 2.09 (SD = 1.08), while optimistic estimates decreased to 0.70 (SD = 0.68). A repeated-measures ANOVA found significant differences among these categories ($F(2, 48) = 15.58$, $p < .001$). Post-hoc pairwise comparisons showed that realistic estimates were significantly higher than both pessimistic ($p < .001$) and optimistic estimates ($p < .001$), with no significant difference between pessimistic and optimistic estimates ($p > .005$). This suggests that participants predominantly made realistic estimates, with no clear preference between pessimistic and optimistic ones, indicating a lack of optimistic planning.

A direct comparison of estimate changes before and after the intervention showed a trend but no significant increase in pessimistic estimates ($t(24) = -2.61$, $p = .015$), no change in realistic estimates ($t(24) = -1.49$, $p = .149$), and a significant decrease in optimistic estimates ($t(24) = 3.87$, $p < .001$). This suggests that the reduction in optimistic bias was primarily driven by a decline in over-optimism rather than a shift toward greater realism or pessimism.

Changes in self-reported time management outcomes

Self-reported questionnaire data showed a significant decrease in trait procrastination from 13.96 (SD = 4.42) to 11.84 (SD = 4.63) ($t(24) = 3.17$, $p = .004$) and a significant increase in general self-efficacy from 18.76 (SD = 2.74) to 19.8 (SD = 4.53) ($t(24) = -3.09$, $p = .005$). Perceived control of time, however, significantly decreased from 20.44 (SD = 3.65) to 12.52 (SD = 3.38) ($t(24) = 10.64$, $p < .001$).

Table 5.1: Summary of Themes and Sub-themes from Thematic Analysis of Study 2 Interview Data

Theme 1:	**Surprising Realizations About Time Expenditure**
Sub-Themes:	Less Time on PhD Work, Shorter Work Day and Week Duration, Overestimating and Underestimating Tasks, Inefficiencies in Time Management
Theme 2:	**Aligning Productivity Expectations With Reality**
Sub-Themes:	Validation and Confidence from the Data, Positive Motivation and Reduced Guilt, Stress from Misaligned Expectations, More Salient External Factors
Theme 3:	**Realism Did Not Always Equate to Accuracy**
Sub-Themes:	Intending to Be Realistic, Acceptable Flexibility of Estimates, Judgments Based on Typical Task Durations, Strategic Pessimism
Theme 4:	**Optimizing Work Habits**
Sub-Themes:	Awareness and Intentionality, Improved Prioritization, Better Break-taking, Improved Focus and Procrastination, Making Peace, Insufficient Reflection

5.8.3 Interview data

We extracted four themes from the interview data that captured the experiences participants had while tracking actual time spent on tasks. Table 5.1 shows a summary of the themes, together with seventeen sub-themes.

Theme 1: Surprising Realizations About Time Expenditure

Tracking increased participants' awareness of how they allocated their time, particularly in relation to their priorities. Some found it helpful for making more realistic assessments: *It was helpful in realistic assessments, it gave me valuable insights into what I'm actually doing during the day, and tasks I should be prioritising* (P23). The tracker also revealed unexpected insights about time spent on PhD work, with many realizing they were dedicating **less time to their PhD** than to other commitments, such as jobs or caregiving: *It's very obvious at the moment I'm overcommitting to my work and under-committing to my PhD* (P14). This often conflicted with their initial assumptions: *I probably do slightly less PhD work than I thought because of having all of these other commitments, and on some level I think I should do more PhD work* (P1). P12 suggested a pie chart visualization to compare time spent on PhD versus other tasks, finding it useful for managing feelings of guilt: *I'd like a pie chart showing time allocated to PhD tasks versus other tasks, as it made me feel guilty initially.*

Participants often perceived their workdays as longer than they actually were, relying on tiredness as a proxy for productivity: *I had no idea at all, I just felt tired and I thought I had worked forever* (P18).

Many felt they worked *all the time* on their PhD, but tracking revealed they averaged about four to five hours of focused work per day. This was often shorter than expected, as the data reflected actual work time rather than the full workday. *When people say, I worked 8 hours, they don't work for 8 hours, they have breaks in between. I would think I've done so little, but actually I felt so tired because I had been intensively working in total for five hours, which spread across the day* (P21). Similar realizations occurred at the weekly level: *I thought I was working 40-60 hours a week but it is actually a lot less* (P5).

Some tasks took longer than expected, while others were shorter. Cognitively demanding tasks, such as thesis writing and summarizing research articles, were often **overestimated**: *I spend less time on research reading tasks than I thought* (P1). In contrast, tasks like preparing or running experiments, admin work, and analyzing data were frequently **underestimated**. For instance, P10 noted: *Running tests in the lab ended up taking much longer because I was doing mistakes because I am so new to this* (P10). Errors and interruptions made time control challenging: *I want to stick to the schedule I set, but I might receive an email that requires two hours of work and it throws the plan out of the window* (P27). Job applications and unfamiliar tasks were also harder to estimate and often took longer than expected.

The tracker data revealed **inefficiencies** in how participants spent their **time**: *I'm probably not as efficient as I would like to be* (P21). Work was often more fragmented than expected: *The tracker showed many little chunks of work and, thinking about it, it's probably more efficient to stick to one thing* (P1). Participants became aware of time-wasting patterns they previously overlooked: *I realised in the beginning that I was wasting a lot of time [...] on emails, especially in the mornings. [...] The tracker really helped me to understand that this is a problem.* (P17). Common time wasters included taking long breaks: *One time, I took a break and it was an hour. I know it's an hour because I was tracking and that was shocking!* (P6), and browsing social media: *I'm able to waste an hour just crawling the Internet* (P3). They also found that travel, cleaning, and preparing for university took more time than expected.

Theme 2: Aligning Productivity Expectations with Reality

The tracker helped participants align their expectations with the reality of their productivity, providing a sense of **validation and confidence** through clear evidence. As P17 shared: *Now, I have proof* *of my productivity! If anyone questions me, I can show them the data.* This evidence also justified their feelings, as P18 noted: *It did justify the tiredness, it's impressive that I still did this amount [of hours] despite that tiredness.* Participants recognized their progress, even when it felt intangible, which was particularly important in the PhD context: *We work for weeks without seeing results, that's a big challenge of doing a PhD* (P12).

Participants reported working with higher **motivation** and a more positive outlook: *This was a positive experience and it made me more motivated to work because I was seeing all this progress* (P2). Feedback provided concrete evidence of their work, helping them realize they were more productive than they thought. As P2 noted: *I thought to myself, actually, you've done more than you thought because the brain is a funny thing, isn't it? Sometimes it tricks you into thinking you were lazy.* Recognizing that their time was well spent, even if not always on priorities, led to satisfaction and **reduced guilt.** P5 shared: *It made me feel less guilty,* while P1 added: *It made me feel less stressed because I was seeing all these tasks getting done.*

Aligning expectations with reality was challenging, with some participants finding feedback **stressful**, especially when tracked hours fell short of expectations. Those expecting 8-hour workdays but managing 5-6 hours adjusted their sense of accomplishment: *Even if I don't work a lot of hours, like four or five, but it's actually good work - I'm more at peace with myself* (P5). Others struggled with ingrained standards: *Even though I realized more than 5-6 hours per day is not possible, I still feel really bad about myself* (P8). Some disliked the tracker because it *made them work in the evenings to record enough hours* (P6). Participants also questioned academic productivity norms, suspecting they underestimated their own work patterns: *I might think that one hour of procrastination with social media a day is terrible, but who knows - it may be better than most other people* (P28).

When external factors, such as additional commitments, challenging environments, or poor supervision, disrupted participants' work, the feedback **made these issues more salient**, often uncomfortably. P3 and P28 expressed frustration and powerlessness when realizing their struggles were beyond their control. For P3, financial pressures took priority over PhD work: *It's not a time management issue... I had no choice but to take extra jobs, and now it's difficult because everyone else has more experience than me.* P28 highlighted a lack of support: *I have not had a positive experience during my PhD [...] my supervisor was indifferent to my challenges [...] There's a structural issue with a lack of support for disruptions like lab closures or sickness.* Timing also played a role, as P25 noted: *The tracker highlighted that I was not productive, but I couldn't do anything about it because I* *was moving houses.*

Theme 3: Realism Did Not Always Equate to Accuracy

Participants noted that whether an estimate was optimistic, realistic, or pessimistic often depended more on **intent** than accuracy. As P19 explained, *I was trying to be realistic... but I haven't done this tons of times before, so I'm guessing how long it might take.* Similarly, P10 estimated 4 hours and 30 minutes but spent 3 hours on the task and 1 hour 30 minutes on an unplanned IT issue, considering the estimate realistic despite the disruption: *I was trying to do the review, but then the IT issues came in.* Even accurate estimates were sometimes labeled pessimistic when extra time was intentionally added as a buffer: *I just gave it a little extra time because I knew how long it might take, and eventually it did take me about an hour* (P21).

Participants varied in how much **flexibility** they allowed in their estimates before considering them biased. Some saw large deviations as realistic, viewing estimates as rough approximations rather than precise predictions. As P20 explained about a 15-minute estimate that took 30 minutes: *It was realistic because I figured it would take me roughly this amount of time.* Others viewed even small differences as signs of bias. P14, who estimated 20 minutes but spent 30, noted: *I was slightly optimistic. I thought it would take me slightly less time than it did.* Many found estimation more useful for tasks requiring accuracy, particularly writing: *If I was writing my thesis or a paper, the tracking would come in very handy to make sure I'm on track* (P12).

Sometimes tasks were judged as realistic because the estimate aligned with the **typical duration** for such tasks, irrespective of the accuracy. For instance, P9 considered her estimate realistic despite its inaccuracy, explaining: *I left myself 30 minutes knowing that I will think about more than just the agenda.* Similarly, P20 stated: *I'd say this is realistic because based on experience, I've done similar tasks and it takes that long.* However, some participants recognized these discrepancies and acknowledged them as an indicator of bias, as P19 reflected: *Maybe I was being optimistic then. Yeah, because it obviously ended up taking longer maybe.*

Participants sometimes adjusted their estimates, with some deliberately using *strategic pessimism* (P17) to counter their tendency toward optimism. Even when an estimate was accurate, it could still be seen as pessimistic if extra time was intentionally built in. P21 considered their correct estimate pessimistic because they had time left for additional work: *Pessimistic - I had to check the* *tracker and make sure it was OK but then it went on to prepare for today's meeting.* Others adopted a pessimistic *just in case* mindset, as P1 noted: *It's probably slightly pessimistic because if something had come up, if you realize through the meeting that someone needs more support, then I think 90 minutes is fine to allocate just in case that happens.*

Theme 4: Optimizing Work Habits

Feedback on time spent on tasks encouraged participants to try new strategies to improve their planning. Some of the effects of the study were increased **awareness** of how they were spending their days, prompting reflection and efforts to be more efficient and focused. P19 noted: *[The tracker] has made me think about being more deliberate in deciding what is important to get done and understanding what my own priorities need to be* (P19). Some participants found overestimation more effective than optimistic goals: *This tracker opened my eyes to the fact that it's better to overestimate than underestimate [...] if I overestimate, I noticed I wasn't likely to watch YouTube, I moved on to the next task* (P5). Others realized they were overestimating tasks unnecessarily: *I definitely think going forward I would shorten the amount of time I would assign to certain tasks* (P20).

Participants began **prioritizing tasks** more effectively. P21 found tracking helpful for perspective: *If it was a bigger, more time-consuming, and difficult task, its importance made me prioritize it over smaller, less important tasks, even if those were easier to do.* Some adjusted their approach by tackling longer tasks first: *If something takes hours, I do it first. In the past, I just listed tasks and started with the first one* (P7). Others prioritized their energy, ensuring key work was done: *It helped me realise I was lacking energy, so I decided to do a little less work and focus on what really mattered.* Some planned fewer tasks to boost productivity: *If you say, right, you've got an hour to do this [...] it makes you a bit more disciplined.... [The tracker] reminded me how I work much better by doing less* (P2).

Participants reported **improving** their **break-taking** behaviours because some of them also kept track of breaks in the trackers: *I think that because I was also keeping track of my breaks, I'm now better at taking a break when I need to and then getting back to work* (P21). A benefit of recording breaks was that participants felt it was easier to return back to work after a break: *I think it definitely made me get back to work quicker. Like right, stop looking at random stuff on your phone.* (P9).

Tracking improved **focus** and reduced procrastination. P14 noted increased concentration: *Some-*

times I needed a break after half an hour, but now I can often work for up to two hours and get a lot done. Some found the tracker made procrastination more visible: *The way I have my tracker here is Friday is pink, Saturday is gray, and Sunday is green. I don't want to mess up the color, so I stick to it* (P13). Higher accountability also played a role: *The days when I was doing the tracker, I didn't really procrastinate. It held me accountable* (P14). This accountability was especially strong at the start of the intervention, as P23 noted: *Because it was the first day of planning, I was thinking a lot about it and being quite detailed. So yeah, I think it looks more accurate.*

Some participants found that tracking helped them *make peace with themselves*, recognizing time management challenges as a natural part of their work. *I have this working habit for years [...] And I don't think it's something that urgently needs to be changed* (P8). Rather than aiming for drastic changes, many appreciated that their productivity was sufficient despite minor inefficiencies. *[The tracker] reduced my stress and gave me a positive feeling because I was getting a lot done even though I was checking Twitter so many times each day – which I should really stop doing* (P1).

Most participants did not take explicit reflection notes, as reflection required significant mental effort, which felt overwhelming alongside their workload: *It takes a lot of effort to stop and think [...] I did not want to face it* (P4). While they recalled *thinking* about their habits, they did not explore them in depth. This **insufficient reflection** led some to express a need for more support: *It would have been helpful to have weekly chats to go over the data* (P14).

5.9 Discussion of Study 6

This study explored how task duration feedback, delivered through a simple manual time-tracking tool, affects both perceived and objective optimistic planning bias, as well as broader productivity outcomes such as procrastination and self-efficacy. It builds on the findings from Study 5 by addressing key barriers to engagement and introducing a more accessible intervention format. Together, these two studies address the third research sub-question of this thesis: *What are the effects of task planning debiasing interventions on planning and productivity outcomes?*

The study shows that manual time tracking can increase awareness of overly optimistic plans (as indicated in interviews) and reduce *perceived* over-optimism in daily planning. Participants realized that setting overly ambitious goals led to procrastination, while more achievable goals reduced this tendency. As several participants noted, *they worked better by doing less*. This finding suggests that recognizing productivity limits does not necessarily demotivate individuals; instead, participants felt more energized by accomplishing realistic goals. Even without mandatory daily planning, many engaged in it naturally, possibly benefiting from more attainable expectations. This aligns with the *goal gradient effect*, where motivation increases as individuals perceive themselves nearing completion [147], and with goal-setting theories, which emphasize that goals should be challenging yet achievable to sustain motivation [25].

Interestingly, the study found that **objective optimistic bias**, measured as the discrepancy between estimated and actual task durations, emerged in the second set of diaries but not in the first, with an average increase of 26 minutes. In other words, participants accurately estimated task durations on the first day of tracking but later overestimated how long they would work. This contradicts prior research, which typically finds overestimation bias to be widespread [77]. One possible explanation, based on participant accounts, is that they were more engaged with the intervention at the start, ensuring their estimates matched actual durations, but paid less attention as the study progressed (e.g., as P23 noted). Since prior research finds optimism bias in observational studies rather than in interventions where participants actively attempt to improve their habits, it is reasonable to conclude that initial motivation influenced this result. The key question, then, is why objective bias did not improve by the end of the intervention, even though participants *perceived* their optimism in planning had decreased.

A plausible explanation is that participants planned **less ambitious tasks** in the second diary, allowing them to complete them more effectively, making the discrepancy feel less problematic. For example, planning to read one article in two hours versus five articles in four hours creates different perceptions of realism. Completing a smaller goal, such as reading one article in 1 hour and 30 minutes, may feel more accurate, whereas failing to read five within four hours could seem overly optimistic and demotivating. This aligns with participants' accounts that accuracy alone did not determine realism in planning. Future studies could control for factors like task completion to further investigate this effect. Additionally, since most participants (20 out of 28) started tracking on a Monday, their final diary entries typically fell on a Friday. This could explain why they encountered more unplanned work later in the week, prompting them to re-organize their schedules and complete planned tasks more quickly.

5.9.1 How can task duration feedback reduce the optimistic planning bias?

From participants' perspectives, realism in planning was more flexible and nuanced than prior work suggests, which often defines it through narrow measures of absolute accuracy [60]. As discussed earlier, task duration feedback improved *perceived* bias while *objective* bias worsened. This divergence highlights the need to consider subjective bias alongside objective measures, especially if improvements in subjective perception contribute to productivity outcomes, such as reduced procrastination and increased self-efficacy. If subjective measures align with meaningful behavioral changes, they warrant greater attention.

By examining *perceived* bias and participants' reflections, we found that the reduction in perceived bias stemmed primarily from participants adjusting overly optimistic estimates, rather than increasing realistic ones or decreasing pessimistic ones. Combined with participants' insights that overloaded to-do lists hinder productivity, these findings suggest that the key challenge is not eliminating minor inaccuracies in time estimates but mitigating overly ambitious plans that become unachievable. Building on Kim et al.'s [148] exploration of time management in time-aware systems, we extend the concepts of punctuality and efficiency. Our findings suggest that punctuality, aligning actions with planned schedules, is less about rigid adherence to every planned task and more about selectively addressing the most unrealistic estimates that threaten productivity. In turn, this approach can improve efficiency, defined as completing tasks with minimal time and effort while achieving desired outcomes.

Some PTM tools, like Sunsama, prompt users to reflect on whether their workload is achievable, but these prompts are often too general (e.g., simple suggestions to move tasks to the next day), offering little guidance on which aspects of a plan need adjustment. A more effective approach would be to unobtrusively detect and address the most overly optimistic estimates rather than requiring users to review their entire plan. Flagging tasks with significant discrepancies, such as those frequently rescheduled or causing delays, could provide more actionable feedback. Additionally, incorporating flexible deadlines, such as "by the end of the week," as suggested by [9], could help users better manage daily fluctuations while maintaining overall progress.

5.9.2 When is task duration feedback beneficial?

Feedback was beneficial to participants when it led to actionable insights regarding their productivity. The most common benefits were observed when the feedback helped participants identify realistic expectations for how long a task would likely take to complete, giving them a clearer understanding of their capabilities. This clarity often triggered what Cox et al. referred to as *digital epiphanies*-moments of realization where participants either adjusted their behaviors (*change* epiphanies) or accepted the need to revise their expectations (*acceptance* epiphanies) [75].

Change epiphanies occurred when participants realized that they were falling short of their goals due to factors within their control, such as procrastination, distractions, or over-optimistic planning. These realizations, while sometimes stressful at first, were often motivating. The increase in awareness is likely linked to the decrease in perceived control over time. Participants became more conscious of their behaviors and took steps to adjust their work routines, such as prioritizing tasks or limiting distractions. This process of change likely improved their self-efficacy and reduced trait procrastination. This finding aligns with the work by Kim et al. [30] that shows how feedback on distracting activities decreases the amount of time engaged with them. In contrast, acceptance epiphanies arose when participants realized that certain obstacles, such as external constraints or workload fragmentation, were beyond their control. In these cases, participants adjusted their expectations and *made peace with themselves* without necessarily changing their habits. They accepted that their productivity might be lower than expected but were still on track to meet their long-term goals.

The emotional impact of these epiphanies, however, was context-dependent. When participants felt that they could address the issues revealed by the feedback, the experience was perceived as beneficial. Conversely, in situations where systemic barriers (such as poor supervisory relationship) or overwhelming workloads (such as doing a PhD, a side job and moving houses at the same time) made change difficult, the feedback sometimes led to stress and frustration. This suggests that the perceived ability to act on the feedback plays a key role in determining whether the *epiphany* is beneficial. This finding extends prior work by Kim et al. [30] by suggesting that the perceived ability to act is a key factor to consider in feedback interventions. These findings suggest that instead of binary metrics like *task done*, tools could track *time invested* or *knowledge gained*, prompting users to reflect on progress made. Additionally, tools could use supportive language and features that support self-compassion, especially when significant gaps between expectations and performance are revealed and the perceived ability to act on the feedback is low.

5.9.3 In what ways is task duration feedback beneficial?

The study suggests that task duration feedback was beneficial to participants when it was targeted at specific planning contexts, fitted naturally into their routine and was delivered through simple tools.

By being tailored to specific contexts

By tailoring feedback to their specific work contexts, participants could connect insights to their unique habits and priorities, making the information more relevant and actionable. Feedback was particularly useful when it helped them understand time distribution across PhD work, other commitments, self-care, or total hours dedicated to a project. Many proactively customized their spreadsheets to track these areas of interest. This finding aligns with prior research on personal informatics. Kersten-van Dijk et al. [149] reviewed empirical studies on personal informatics systems and found that users benefit most from actionable, context-specific insights. Similarly, Huang et al. [150] showed that on-calendar visualizations helped users relate fitness data to their daily schedules,which made the insights more actionable and relevant.

Despite this evidence, prior research indicates that PTM tools fail to adequately help users obtain tailored feedback. Tools lack support for "custom analytics" defined as features enabling users to visualize and analyze time tracking data tailored to specific tasks and activities [100]. The apps primarily supported visualizations of task management data, such as number of tasks done. Custom analytics requires users to label tasks with categories for the analytics to work, but a study found that labeling time tracked data over a period of time was not supported [100]. This limitation aligns with a key barrier to reflection in personal informatics tools-poor visualizations of time data-identified by Li et al. [151]. These challenges may also contribute to low engagement with tracking itself; without intuitive tracking, there is little meaningful data to generate actionable feedback, keeping demand for advanced analytics low. Addressing these gaps requires both simplifying the tracking process and exploring new ways to implement tailored time-tracking feedback.

By being collected and delivered as a part of routine

By integrating time tracking into daily routines, Study 6's approach-logging tasks at natural break points-proved more effective in sustaining the habit of tracking compared to Study 5's start-and-stop timer method. Prior research supports this, showing that habit formation is more successful when tracking aligns with routine behaviors [126]. However, many popular time-tracking tools, such as Toggl and TimeDoctor, rely on continuous task-based tracking, which may not align with users' routines. To improve, tools could incorporate **flexible tracking methods**, such as proactive prompts to log time after breaks or end-of-day reflections to categorize time into broader activities (e.g., PhD work or side jobs). Playback features, like ManicTime's timeline or ScreenTrack [152], could provide visual summaries to help users review their work patterns. Extending this idea, incorporating diary-style tracking where users highlight activities of interest could reduce the burden of tracking while maintaining meaningful insights, aligning with Choe et al.'s [153] work on semi-automatic tracking in planning contexts.

By being delivered through simple tools

Simpler tools, like the spreadsheet tracker in Study 2, may be more effective for delivering time-use feedback than complex PTM apps, particularly for non-expert users. Reducing intervention complexity appeared to help participants focus on task durations, making this information more salient [135]. Offering too many features could sometimes hinder rather than enhance productivity, as extensive functionalities may overwhelm users and reduce engagement [100, 154]. Users might benefit more from streamlined tools that align with their experience levels and immediate goals.

PTM apps could improve by integrating coaching functionalities before introducing advanced features. Currently, they differentiate basic and premium versions by pricing rather than guiding users through progressive feature adoption. Their focus on maximizing productivity-often framed as achieving more in less time-may unintentionally reinforce planning biases. Helping users set fewer, more intentional goals could better support productivity by shifting the focus from efficiency to fostering self-knowledge and work-life balance.

5.9.4 Limitations specific to Study 6

First, the lack of a control group, though typical in exploratory research, restricts causal conclusions about the intervention's impact on planning bias. Notably, objective bias worsened, suggesting the intervention may have contributed to this outcome. However, the absence of a control group leaves uncertainty about whether this was due to higher effort on the first day of tracking. Additionally, uncontrolled factors such as participants' initial motivation, the study mostly beginning on Mondays and ending on Fridays and the lack of measures for task type or difficulty may have influenced the results. Second, the lack of guided reflection on tracked data may have reduced the intervention's effectiveness. Third, a limitation in the design was that participants assessed the bias in both diaries after the intervention which may have led to a recall bias as the second diary was more recent and easier to remember. This decision was taken to avoid revealing the main study goal to increase the realistic estimations early on. Finally, while a custom-made tool may have been more tailored to the study, using an evidence-based commercial app provided valuable insights into the practical effectiveness of existing tools.

5.10 Summary of Chapter 5

Chapter 5 reported two field interventions that examined task duration feedback as a debiasing strategy in academic work. Study 5 (N = 10) trialled a commercial PTM application (Sunsama), revealing limited engagement and highlighting design barriers that prevented feedback from being embedded into daily practice. Study 6 (N = 30) used a simpler spreadsheet-based tracker, which achieved higher engagement and reduced perceptions of over-optimism, though without measurable improvements in objective accuracy. Across both studies, participants showed diverse interpretations of what counts as realistic planning, emphasising task type, intentions, and personal capacity rather than accuracy alone. Together, the findings demonstrate that duration feedback can support planning when delivered in lightweight, flexible formats, but its effectiveness depends on how users interpret and integrate feedback into their routines. This chapter contributes to HCI by showing both the potential and limitations of debiasing interventions in real-world contexts and by drawing design implications for PTM tools that align with varied user needs.

Chapter 6

General Discussion

Chapter Outline *This chapter provides a synthesis of findings, highlights contributions to theory and design, reflects on the methodological approach, discusses practical implications, and outlines limitations and directions for future work.*

This chapter brings together findings from the six studies to show how optimistic planning bias manifested in academic work. Across the studies, optimistic planning bias appeared in three main ways: systematic misestimations that varied by task and context, disengagement from planning when overruns accumulated, and continued reliance on planning for structure and motivation despite some estimation inaccuracies. The intervention studies demonstrated that task duration feedback did not reliably improve minute-level accuracy, but diary and interview data indicated shifts in how participants classified their plans, reported reductions in procrastination, and adapted their planning practices. The theoretical contribution developed from these findings is the concept of reflective planning, which describes how people respond to mismatches between plans and outcomes by adjusting, reframing, or occasionally disengaging from their plans. This perspective highlights that technologies should not focus only on correcting estimation errors but should also scaffold the processes that help users sustain planning in practice. Examples include support for task breakdown and prompts to reconsider commitments when schedules are missed. The work is limited by its focus on academics, the limited duration of intervention trials, and the exploratory nature of the studies. Future research should explore reflective planning in diverse knowledge work domains, test its long-term effects, and assess how new tools, such as AI, can scale reflective planning support.

6.1 Answering the research questions

This thesis set out to investigate how technologies can better support accurate planning in academic work. To address this overarching aim, three research questions were posed. The first examined the challenges academics face in planning tasks accurately, the second considered the extent to which current personal task management tools address these challenges, and the third tested the effects of debiasing interventions in practice. Together, the studies provide evidence that planning inaccuracies are pervasive yet patterned; that existing tools offer only partial support for improving planning; and that debiasing interventions can support reflection, helping people bring plans and outcomes into closer alignment even when absolute accuracy does not consistently improve.

RQ1: What are the main challenges that academic knowledge workers face in planning work tasks accurately? Studies 1 and 2 showed that academics frequently misestimated task durations, with day-level plans often over-scheduled and task-level estimates subject to both under- and over-estimation. Sources of inaccuracy included vague task definitions, hidden sub-tasks, interruptions, and mental fatigue. Study 6, while finding baseline estimates to be relatively accurate on average, revealed that participants still held optimistic expectations about overall progress, leading them to plan beyond what was realistically achievable. Overall, the key challenge with accurate plans was optimistic planning bias: people consistently believed they could achieve more within a given time frame than proved feasible, with patterns shaped by task type and situational pressures.

RQ2: To what extent do existing planning technologies support users in making accurate plans? The psychology review (Study 3) identified four strategies with potential to mitigate optimistic planning bias through technology design: duration feedback, distributional data, task breakdown, and induced neutrality. The functionality review (Study 4) showed that current personal task management (PTM) tools provide only partial support for these strategies. Duration feedback and task breakdown were present, but usually in limited or burdensome forms, while distributional data and induced neutrality were absent altogether. Taken together, this indicates that existing planning technologies do not adequately support users in making more accurate plans.

RQ3: What are the effects of task planning debiasing interventions on planning and productivity outcomes? The intervention trials (Studies 5 and 6) tested task duration feedback delivered through two different formats: a commercial PTM app with debiasing potential (Study 5) and a manual spreadsheet tracker (Study 6). Neither intervention produced reliable group-level improvements in minute-level estimation accuracy. In Study 6, however, diary data showed a shift in participants' perceptions of accuracy, with more plans judged as realistic and fewer as optimistic after the intervention. Interviews suggested that some participants also adapted their planning practices, for instance by breaking tasks into smaller units or scaling commitments, which reduced the scale of deviations even if this did not translate into significant accuracy gains in the quantitative measures. Participants reported subjective benefits, including increased awareness of time use, reduced procrastination, higher self-efficacy, and a stronger sense that plans were realistic enough. At the same time, these benefits depended on context: the supports were most valuable when participants had sufficient control over their work to adjust tasks accordingly, whereas in more constrained situations, improving precision alone offered limited benefit.

By addressing these questions, this thesis offers the following *answer* to the overarching research question: technologies can best support accurate planning by applying debiasing strategies (such as task-duration feedback and task breakdown) selectively to the most consequential inaccuracies, rather than to every minor deviation. In doing so, they scaffold the reflective processes through which people reconcile plans with outcomes. Support would benefit from prioritizing the perception that plans are realistic and achievable, helping individuals form grounded expectations of their productivity, rather than enforcing strict *objective* precision, which is rarely attainable in the lived realities of daily work. Finally, effective supports must be context-sensitive: improved accuracy is most valuable when individuals have the autonomy to act on revised estimates, whereas an overemphasis on time precision in less supportive contexts risks unintended negative effects.

6.2 Synthesis of findings across studies

The findings contribute a broader account of the *lived experience* of optimistic planning bias in academic work. This matters directly for the main question of how technology can better support accurate planning, because without understanding how bias is embedded in everyday planning practices, tools risk focusing narrowly on error correction rather than supporting the reflective and motivational processes through which people actually manage their plans. The cross-study synthesis that follows develops this contribution by organising the evidence into three dimensions of how optimistic planning bias was experienced in practice: patterned misestimation, emotional consequences, and motivational benefits beyond accuracy.

First, patterned misestimation characterised academic planning across the studies. Participants frequently misjudged the time required for work, but the form this took varied by individual and by context. At the level of the working day, many participants in Studies 1 and 5 assumed they could complete more tasks than proved possible, leading to over-scheduling, while others showed estimates that were closer to what was achieved. In Study 6, baseline estimates were relatively accurate on average, with little measurable bias before the intervention, although interviews suggested that participants still carried optimistic expectations about progress. This points to the fact that misestimation was not inevitable, but contingent on differences in planning style, task definitions, situational pressures, and the influence of the study context itself. At the level of individual tasks, patterns were similarly mixed: durations were often underestimated but could also be overestimated, depending on task type and circumstances. Study 1 highlighted sources of error such as vague task definitions, hidden sub-tasks, and interruptions. Studies 5 and 6 showed that these issues recurred in broader field trials, suggesting they were not limited to individual accounts but persisted across daily practice.

These findings broadly align with psychological research on the planning fallacy, which shows that people often underestimate long or complex tasks while overestimating shorter ones [60]. At the same time, they complicate classic accounts of the planning fallacy that treat underestimation as a universal bias, by revealing conditions under which estimates were accurate or even pessimistic [7]. They extend HCI studies that quantified interruptions and task fragmentation as execution costs [12], by showing that vague task formulations and hidden subtasks also function as reasons for planning delays.

Second, emotional consequences accompanied the practical difficulties of optimistic planning bias. When planned tasks overran, participants reported negative reactions such as frustration and discouragement (Studies 1 and 2). In some cases, repeated failures led to disengagement from planning routines (Study 2). In the field trials, similar effects were noted: in Study 5, participants described the discouraging impact of daily overruns, and in Study 6, even when baseline estimates were relatively calibrated, the effort of tracking overruns was sometimes experienced as disheartening. At the same time, participants developed coping practices, such as breaking down tasks or manually tracking time (Studies 2 and 6), which helped them regain a sense of control. These findings indicate that optimistic planning bias was not only a cognitive phenomenon but also an affective one: planning inaccuracies generated strain that could undermine engagement with planning unless counteracted by compensatory strategies.

These findings align with psychological work showing how unmet expectations undermine confidence and persistence. Steel [155] links repeated failures to falling self-efficacy and procrastination, a cycle also visible in our field studies. They also extend HCI accounts of technology breakdowns [9], showing a similar pattern in planning routines, where repeated overruns risk disengagement unless buffered by reflective supports.

Third, the studies showed that planning practices and interventions offered motivational benefits even when accuracy did not improve. In Study 1, participants valued the structure and direction of planning despite frequent overruns, and in Study 2, routines helped maintain momentum during lockdown. The intervention trials reinforced this point: in Studies 5 and 6, task duration feedback did not reliably increase group-level accuracy, but participants still reported greater confidence, reduced procrastination, and a stronger sense of control. Diary data from Study 6 also showed a shift towards classifying plans as more realistic. These findings contrast with psychological experiments (Study 3) and commercial PTM tools (Study 4), which have typically treated accuracy as the main outcome. In practice, the value of planning and debiasing features lay in sustaining motivation and engagement under conditions of bias.

Even modest accuracy gains, and awareness of which inaccuracies mattered most, proved motivating and enhanced participants' sense of control. This suggests a refinement to popular planning frameworks (e.g. [66]): while precision need not be the primary aim, selective calibration of impactful inaccuracies can play an important motivational role and merits more attention in both theory and design.

Overall, the six studies show that optimistic planning bias is not just an occasional mistake but something woven into the fabric of academic working life. It surfaced across very different tasks and contexts, shaping how participants organised their days and touching their motivation and wellbeing. Planning was at once unreliable and indispensable: people misjudged time again and again, yet planning still gave them a sense of order, helped sustain effort, and offered direction. Looking across psychology and commercial tools, it is striking that existing approaches tend to spotlight the moments when plans fall apart, rather than offering support for the ongoing, everyday work of planning. This thesis characterises optimistic planning bias as a patterned, persistent feature of academic work, with consequences not only for accuracy but also for how people actually handle their workday deciding what gets done, what gets postponed, and how to cope with a workload that is always a little too big.

6.3 Contributions to theory

This thesis contributes to theory primarily within Human-Computer Interaction, where planning and productivity tools are designed and evaluated. HCI research has developed theories and concepts for understanding technologies of self-regulation, such as personal informatics [151] and lived informatics [41], and has examined reflection as a process supported by tracking and feedback [156]. Yet planning technologies themselves have rarely been theorised beyond their functional role as organisers of tasks [9, 6]. The findings of this thesis extend this space by showing that planning is not only a matter of scheduling work but also a practice through which people sustain motivation, regulate emotions, and construct a narrative of progress. Secondarily, the thesis contributes to psychology, where optimistic planning bias has typically been operationalised in terms of the statistical accuracy of task duration estimates. While psychological accounts acknowledge cognitive, motivational, and social mechanisms [81], interventions are generally assessed by whether they reduce estimation error. The findings of this thesis challenge this emphasis by showing that bias is lived not only as misestimation but as a feature of everyday practices of organising work.

6.3.1 Reflective planning

This thesis advances the idea of reflective planning as a construct that links empirical observations of bias with design opportunities for planning tools. Across six studies, the evidence shows that planning was not only about predicting task durations but about engaging reflectively with the inevitable mismatches between intentions and outcomes. This resonates with established accounts of reflection [157] as often triggered by breakdowns: moments when an expected trajectory falters and users are prompted to re-examine their approach. In observational studies, this was seen in how academics adapted or re-framed their plans when overruns occurred, but also disengaged from planning altogether when opportunities for reflection were scarce. In review studies, it appeared as a gap in psychological and technological framings, which foreground accuracy while neglecting lived engagement. In intervention studies, it was evident in the ways duration feedback and scaffolds prompted participants to reconsider, rescale, and sustain their planning practices, even without reliable accuracy gains. The cross-study synthesis (above) highlighted these dynamics in three dimensions: patterned misestimation, emotional consequences, and motivational benefits. Together, these findings demonstrate that planning itself functions as a reflective practice, and tools can support this process by scaffolding interpretation and adjustment rather than only enforcing precision.

Reflective planning is introduced here as a *designable construct* situated between specific design features and broad theories of reflection. It extends the PTM literature by making explicit a dimension of planning that has remained implicit in HCI work on productivity tools: people do not simply use plans as forecasts to be validated, but engage with them in selective and interpretive ways [4, 6]. This thesis adds to that literature by specifying the reflective dimension of those practices, how plans are interpreted, suspended, or re-purposed, aligning with long-standing CSCW and HCI arguments that plans are resources for action, not scripts [80]. While narrower in scope than established sensitising and strong concepts such as seamfulness or social navigation, reflective planning can be understood as a form of intermediate-level knowledge: empirically grounded, generative for design, and transferable across settings without aspiring to the generality of formal theory. [158].

Three features distinguish reflective planning. First, it foregrounds the motivational and affective roles of plans: participants relied on plans not only to schedule tasks but also to restore a sense of control and maintain confidence in their ablities at work. Second, it recognises the selective and situated nature of planning: individuals adjusted how much detail they recorded, when they estimated, and how strictly they adhered to plans, depending on context. Third, it identifies planning itself as a site of reflective engagement. While reflection in personal informatics is often discussed in relation to retrospective data, here it took place prospectively through the very act of estimating and scheduling. Participants used their plans as artefacts for considering alternative futures, gauging whether commitments were realistic, and negotiating what could be carried forward. In these moments, the discrepancy between plan and reality functioned as a kind of breakdown, a productive disruption that opened space for reflection and adjustment [157].

Self-tracking and planning are closely related in how they produce artefacts that mediate reflection. In personal informatics, Rooksby et al. [41] showed that self-tracking was rarely about generating comprehensive or precise datasets. Instead, people engaged in what they called lived informatics: appropriating tools selectively, dropping and resuming tracking as circumstances changed, and interpreting incomplete data in relation to immediate concerns. Importantly, they observed that tracking was often prospective rather than retrospective, used to orient people towards where they were heading in the short term rather than to assemble accurate historical records.

Planning practices exhibit similar qualities. To-do lists and schedules record intended future behaviour, creating potential discrepancies between what is planned and what is realised. As with tracking, these artefacts were not treated by participants as forecasts to be validated against outcomes, but as flexible resources to be reinterpreted. Across the studies, academics recalibrated expectations, set aside plans that no longer fit, or revisited them for reassurance when progress felt uncertain. At the same time, when opportunities for reflection were scarce, some disengaged from planning altogether, echoing Rooksby et al. [41] finding that people often abandoned or suspended tracking when it no longer felt useful. Reflective planning therefore extends lived informatics, showing how imprecision in planning is not simply tolerated but actively used in negotiating motivation and control in everyday academic work.

Reflective planning offers HCI three concrete affordances. First, as an analytic lens. It provides a vocabulary for evaluating technologies in terms of how they support interpretive engagement with plans. Rather than asking whether features improve accuracy, researchers can examine whether they create space for rescaling, renegotiating, or reframing commitments. This shifts evaluation criteria from error reduction to the quality of reflective engagement. Second, as a generative principle. It points towards designing lightweight scaffolds that stabilise reflection without imposing rigid accuracy demands. Examples include optional task breakdowns when overruns recur, flexible rescaling of daily schedules, or feedback that frames carry-over as an adjustment opportunity rather than a failure. These principles translate the interventions tested here into actionable design directions (see further in Design Implications below). Third, as a research programme. It raises new questions for HCI: How does reflective planning unfold in collaborative settings where plans are shared? What design approaches enable AI to scaffold reflection rather than optimise away user agency? And what are the long-term effects when opportunities for reflection are absentdoes disengagement accumulate into burnout? These questions extend the scope of reflection research into planning and position reflective planning as a starting point for sustained inquiry (see Future Directions).

6.3.2 Extending psychological accounts of planning bias

Psychological theories of judgment and decision-making have long characterised the planning fallacy as a deviation from statistical accuracy, explained through mechanisms such as neglect of distributional data or biased memory recall [57, 59, 107]. Within this framework, success is typically measured by whether people reduce their estimation error. Prior interventions tested in psychology, such as providing distributional information [59] or task segmentation [69], have been evaluated almost exclusively in terms of whether they improve accuracy. The present research confirms that over-optimism is indeed a pervasive feature of planning, but it also shows that this narrow emphasis on accuracy fails to capture the full significance of planning practices.

Across Studies 5 and 6, participants reported benefits from debiasing strategies even when accuracy did not improve. For example, in Study 6 diary data showed that participants classified a greater proportion of their daily plans as realistic during the intervention, despite no group-level change in accuracy. Participants also described feedback as helpful because it reassured them that their plans were realistic enough and reduced procrastination. These findings highlight the importance of distinguishing between objective accuracy and what can be described as *subjective realism*: the sense that plans are workable and aligned with actual capacity. Unlike accuracy, subjective realism captures the motivational and affective dimensions of planning, which shaped how academics sustained effort and commitment in everyday work. This suggests a useful direction for future psychological research: examining not only the objective accuracy of plans, but also how subjective realism influences motivation, persistence, and wellbeing.

Study 1 developed a typology of ten slow-downs that contributed to optimistic planning bias in academic work, grouped into four themes: preparatory work, breaks, requests, and fatigue. This extends psychological models that have typically located the problem in cognitive mechanisms. The typology highlighted that misestimations often arose from task properties, such as vague task definitions, omitted subtasks, and unplanned preparatory work. Psychology has emphasised novelty and complexity as drivers of underestimation, but typically in controlled lab settings [93]. HCI studies, meanwhile, have observed how vague tasks and hidden subtasks complicate everyday work [159, 160], yet have not framed them as systematic sources of bias. The typology also pointed to the role of contextual conditions, including requests, interruptions, and fatigue. HCI has richly documented phenomena such as multitasking and fragmented attention [12] but these have usually been presented as challenges of knowledge work rather than also as factors in time misestimation of planned tasks. Study 1 reframed them as amplifiers of bias, showing how workload norms, unexpected requests, and the rhythms of fatigue directly shaped the divergence between plans and outcomes.

Together, these contributions extend psychological accounts in two directions: first, by showing that bias matters not only as an error to be corrected but also as part of an affective economy that sustains motivation through subjective realism; and second, by situating its causes in the interaction of cognitive, dispositional, and contextual factors. This broader view integrates prior experimental insights with HCI observations of work challenges, reframing them as bias factors.

6.4 Design implications

Introducing reflective planning as a design construct fills a gap in HCI. Prior work on planning and task tools has focused on capturing tasks (e.g. [1]), supporting task editing and text-editor assistive features (e.g. [26, 117]) or on increasing self-awareness and reflection by limiting daily task load (e.g. [161]). Yet those approaches do not directly tackle optimistic planning bias or repeated deferral in everyday planning. Reflective planning extends this work by focusing on how tools can help users stay engaged with plans even when tasks unfold more slowly or reveal hidden subtasks.

This perspective reframes what counts as effective planning support. The aim is not simply to increase throughput or ensure checklist completion, but to help individuals sustain subjective realism: the ability to maintain motivating yet feasible plans under uncertainty. Design interventions informed by reflective planning can therefore span multiple domains: in PTM tools, they involve detecting oversized today-list items and surfacing deferral trajectories; in collaborative work systems, they mean supporting collective realism through shared feasibility cues; and in emerging AI assistants, they entail moving beyond acceleration toward coaching-like mechanisms that scaffold reframing, negotiation, and meaning. This section concludes by translating these implications into ten concrete recommendations for the designers of PTM and AI task tools.

6.4.1 Implications for personal task management tools

Findings from Studies 1 and 2 showed that many of the difficulties with planning accuracy emerged not from lack of effort but from the way daily intentions were expressed in today lists. Participants often wrote tasks that were too broad (finish article, prepare lecture), only to discover that progress required multiple hidden steps. These oversized items repeatedly rolled forward, leading to discouragement and disengagement. One implication is that PTM tools should scaffold task framing at capture: for instance, flagging vague entries that lack actionable language or tasks repeatedly deferred without decomposition. Current tools such as Todoist, Microsoft To Do, or Asana can warn when scheduled tasks exceed available calendar time, but they rarely diagnose structural vagueness in daily lists. Our contribution is to foreground this overlooked layer by treating the scoping of today lists as a locus of bias, and proposing mechanisms such as breakdown detection to help users recalibrate expectations before discouragement sets in.

These practices highlight that what matters to users is not only whether a task is completed but the path it takes over time. Prior HCI work has noted the importance of flexible capture environments [26] and lightweight representations of progress (e.g., Baumer's reflective informatics [157]). Building on this, the findings point to the need for tools that make rollover trajectories visible and interpretable. Rather than presenting repeated deferral as failure, systems could embed lightweight annotations (e.g., draft written, needs editing), visual histories of rescheduling, and prompts to record why a task is rolling over. By surfacing whether deferrals reflect partial progress, shifting priorities, or scope creep, planners could support users in sustaining motivation and feasibility.

This perspective also reframes daily plans themselves. Participants valued them less as predictive devices than as affective scaffolds: laying out a plan offered order and control even when execution diverged. Lund and Wiese [26] showed how planning can combine productivity with emotional regulation; our findings emphasise that this affective dimension becomes especially critical under repeated rollover. The implication is that digital planners should not be judged only by whether plans come true, but by whether they help individuals remain engaged. Features that acknowledge partial achievement and allow reinterpretation would align digital systems with longstanding practices in paper planning, where users flexibly rework intentions to balance aspiration with reality [162].

Finally, LLM-based tools offer new ways to extend these affordances. As Studies 2 and 5 showed, users benefited when they were able to reinterpret repeatedly deferred tasks. LLM-enabled tools could identify such patterns and propose constructive reframing, suggesting breakdown, priority renegotiation, or postponement. Whereas earlier prototypes have emphasised flexibility at capture, our contribution is to demonstrate how AI companions might operationalise reflective scaffolds: highlighting overloaded today lists, simulating trade-offs, and reframing rollover as subjective realism rather than failure. This could help shift PTM tools from passive recorders to active companions that help sustain motivation and realism, reducing the risk of disengagement and burnout, if such tools are implemented in human-centered ways [163].

6.4.2 Implications for general work tools

Collaboration platforms such as Slack, Teams, and project trackers like Trello or Asana have been widely studied in HCI as coordination technologies that support accountability, visibility, and throughput [164]. Yet findings from this thesis suggest that accountability alone can be corrosive when it encourages individuals and teams to ignore feasibility. Across Studies 1 to 4, repeated deferral and chronic overcommitment were not merely individual failings but symptoms of systemic overload. Current platforms often exacerbate this problem by surfacing completed items, missed deadlines, or pending counts without contextualising whether workloads were realistic in the first place.

The implication is that collaborative tools should embed scaffolds for collective realism rather than focusing narrowly on throughput. For example, automated stand-ups could include prompts not only about what was done, but also about whether the plan felt realistic, or whether repeated rescheduling is signalling structural constraints. Instead of treating deferral as slippage, systems could visualise patterns of overload across the team, helping groups recalibrate expectations together. In this way, accountability would be broadened to include responsibility for setting realistic plans, not just for delivering against them.

A similar argument applies to dashboards and email clients. Current efficiency logics encourage users to chase inbox zero or maximise task clearance [165]. Prior HCI research has shown how such framings intensify pressure without helping users manage demands [166]. Findings from this thesis extend this critique: by foregrounding counts and completions, institutional dashboards reinforce unrealistic norms and hide the ecological factors shaping overcommitment. A more constructive orientation would be to highlight distribution of work over time or balance across roles. Visualisations could show how workloads accumulate, how often tasks are rescheduled, or how certain roles consistently absorb more work.

6.4.3 Implications for emerging AI systems

Generative AI systems such as Microsoft 365 Copilot, Notion AI, and ChatGPT are increasingly positioned as productivity accelerators: they draft documents, compress information, and automate scheduling. While powerful, this throughput framing risks deepening the very problems identified in this thesis. If AI is used solely to accelerate work, it may compound optimistic planning bias by enabling people to take on even more than is feasible.

The studies in this thesis point toward an alternative design orientation: AI should be developed not only to accelerate output, but also to scaffold reflection. For instance, an AI assistant embedded in an email client could detect when new commitments risk overwhelming available time, surfacing a reflection prompt before overcommitment occurs. A scheduling assistant could highlight whether a proposed plan leaves adequate slack for unexpected contingencies, drawing on patterns of prior delays. Rather than silently optimising for maximum density, AI could function as a provocateur that interrupts unrealistic expectations and invites recalibration.

This orientation extends earlier HCI work that positioned technologies as reflective provocateurs rather than neutral tools [167]. The contribution of this thesis is to bring that perspective into the rapidly evolving domain of generative AI productivity systems. Embedding reflective provocations into AI systems would broaden their role: from accelerators of throughput to partners in sustaining realistic and motivating relationships with work. Such tools could help individuals resist institutional and cultural pressures to overcommit, counteracting stress and burnout at their source.

By surfacing patterns of unrealistic density, demoralising expectations, or repeated deferral, AI could make visible the hidden costs of optimistic bias at scale. In doing so, it would not only support individual self-regulation but also provide organisations with data to identify when systemic conditions (not individual failings) are driving planning struggles. In this way, generative AI systems could play a role in shifting workplace cultures toward healthier norms of planning and commitment.

6.4.4 Practice-oriented recommendations for developers of PTM and AI task tools

While the primary contributions of this thesis are theoretical and empirical, it is also important to translate the findings, including the above design implications, into actionable insights for those building the next generation of personal task management (PTM) tools and AI-based task agents. Startups and product teams often work in terms of features, adoption, and user engagement, so the following recommendations are expressed in practical language that can guide design and development. The following **ten practice-oriented recommendations** draw directly on the empirical findings across Studies 1 - 6. These recommendations translate the thesis findings into product-level lessons that can inform concrete design choices.

- **Flag big, vague tasks.** Detect monster tasks such as *"write paper"* or *"fix website"* and nudge users to break them into smaller, more actionable steps.
- **Track and surface rollovers.** Show when a task has been repeatedly postponed (e.g., this task has been rescheduled 4 times) and turn that into a reflection point to reframe, break down, or drop the task.

- **Check feasibility, not perfection.** Instead of chasing perfect time estimates, highlight when a daily plan is overloaded (e.g., This looks like a 10-hour day, would you like to reprioritize?).
- **Give light-touch feedback on time use.** Provide quick comparisons of estimated vs. actual time, delivered as simple summaries or nudges, without requiring heavy manual time logging.
- **Make reflection easy.** Allow lightweight annotations (e.g., this always takes longer than expected) and surface these notes back to users in future planning sessions.
- **Keep AI suggestions optional.** Let AI propose task breakdowns, reschedules, or deadline shifts, but ensure that users retain final control with a simple accept/reject action.
- **Show how tasks fit into the bigger picture.** Link daily task lists to weekly capacity, energy levels, or team priorities with simple dashboards (e.g., You are at 80% of your weekly load).
- **Integrate where people already work.** Embed nudges and reflections into calendars, email, or collaboration platforms like Slack/Teams, rather than requiring users to adopt a standalone app.
- **Celebrate progress, not just completion.** Reinforce momentum by acknowledging partial progress or drafts (e.g., 3/5 steps completed. Great progress!) instead of only marking tasks done or not done.
- **Reframe from speed to sustainability.** Position the product as supporting long-term productivity and wellbeing, moving away from "inbox zero" or "clear more tasks" efficiency logics.

6.5 Methodological reflections

Beyond its contributions, this programme of research offers methodological insights into how planning bias can be studied in HCI. Most prior work on the planning fallacy has relied on laboratory experiments designed to measure accuracy under controlled conditions. By contrast, this thesis combined observational studies, short-term interventions, and reviews of existing tools to examine how over-optimism unfolds in the everyday realities of academic work. This approach did not yield uniform measures of accuracy, but it revealed the situated character of planning: how estimates are shaped by interruptions, institutional pressures, and the wider tool ecology in which planning takes place.

A strength of this programme lies in its cumulative design logic. Early qualitative accounts of task estimation and deferral informed later interventions, while intervention findings were interpreted in light of these lived experiences. Reviews of digital tools then helped situate both within the broader technological landscape. This layering across methods enabled triangulation, so that even modest-scale studies could build toward a richer picture when considered together. It illustrates how mixed-methods designs can be used not simply for redundancy but for complementarity, with each method highlighting dimensions of the problem the others could not capture.

Another methodological contribution is the demonstration of lightweight, ecological interventions. Instead of aiming for large-scale behavioural change, the studies used short probes embedded in everyday tools: apps, spreadsheets, digital pens, as a way to test feasibility and capture immediate user experience. This strategy provided insight into which debiasing strategies felt workable and motivating, even if the trials were too brief to establish long-term effects. It shows how small, pragmatic interventions can serve as design probes that generate knowledge about both opportunities and barriers for technology design.

Finally, the programme highlights the value of translation as a methodological move. Study 3 did not attempt a full meta-analysis of psychological evidence, but deliberately reinterpreted experimental findings into concepts usable for HCI design. This translational stance complements traditional evidence synthesis by prioritising accessibility and applicability across disciplines. It illustrates how methodological pluralism, spanning translation, probing, and ecological observation, can advance understanding of a bias that is at once cognitive, technological, and cultural.

In sum, the methodological contribution of this thesis is to show that planning bias can be investigated not only through accuracy metrics in the lab, but through a multi-layered research strategy that attends to lived experience, technological mediation, and design implications. This orientation foregrounds ecological validity and design relevance, offering an alternative model for how HCI can engage with psychological constructs in ways that respect their complexity in practice.

6.6 Practical and societal implications

The findings of this thesis have implications across four domains: universities, workplace wellbeing initiatives, digital mental health, and policy.

6.6.1 Implications for universities

The evidence presented shows that optimistic planning bias is not simply an individual failing but a systematic tendency in everyday work, compounding stress and disengagement. Workload policies and support systems should therefore recognise the tendency to underestimate cognitively demanding tasks such as writing, reviewing, and teaching preparation. Universities frequently encourage staff and students to adopt digital planning tools [11], yet the functionality review (Study 4) showed that widely used systems rarely address planning bias. Institutions should be cautious about endorsing tools that emphasise task organisation alone, and instead support or develop tools that embed reflective features such as duration feedback or task breakdowns, which this thesis found both usable and beneficial.

The findings also speak to the design of wellbeing and professional development initiatives. Time-management workshops, mentoring, and supervisor training could integrate reflective planning practices, particularly for doctoral researchers and early-career academics, where unrealistic planning is a common source of stress. Taken together, these steps suggest that universities can play a more active role in addressing the structural and cultural drivers of over-optimism in planning, rather than placing responsibility solely on individuals.

6.6.2 Implications for workplaces

In knowledge-intensive sectors such as academia, law, and healthcare, planning failures undermine wellbeing. Organisational initiatives often focus on resilience training, mindfulness apps, or performance dashboards, addressing symptoms rather than root causes [168]. This thesis shows that recurrent overcommitment and mismatched expectations are key drivers of demoralisation. Addressing these requires cultural as well as technical change: building reflective planning into everyday practices, from team check-ins to appraisal conversations. The broader stake is whether organisations continue to frame burnout as an individual failure of resilience or acknowledge it as a systemic issue tied to how work is structured and planned.

6.6.3 Implications for digital mental health

Platforms such as Headspace, Woebot, and Calm have successfully scaled mindfulness and CBT techniques, but rarely address planning despite its central role in stress and procrastination. The interventions in this thesis show that even lightweight planning supports for instance, reflecting on expected vs. actual durations can reduce procrastination and improve self-efficacy. Integrating such features into mental health apps could expand their scope from stress management to stress prevention, marking a shift towards upstream mental health support that targets everyday practices rather than only coping mechanisms.

6.6.4 Implications for policy

While individual tools can scaffold reflective planning, their impact is constrained in environments that systematically demand overcommitment. The academic settings studied here exemplify how unrealistic workloads and cultural norms of constant productivity undermine individual strategies. At a policy level, this highlights the need to address structural contributors to planning failure. Universities and other knowledge-intensive organisations could use aggregated planning data not just for individual feedback but to reveal systemic overload. Such evidence could inform workload reforms, shifting planning technologies from instruments of self-management to tools of accountability and empowerment, helping individuals and groups challenge unsustainable environments.

6.7 Limitations

Like all empirical research, the work presented in this thesis has limitations that should be acknowledged. Context and generalisability. The studies were conducted primarily in academic settings, with participants drawn largely from student and researcher populations. While academia is a paradigmatic example of knowledge work, with high autonomy, competing demands, and significant self-management, it is not representative of all professional domains. Knowledge work in industry, healthcare, or creative practice may differ in its rhythms, cultural expectations, and available support. This raises questions about how far the findings generalise. The decision to focus on academic contexts was deliberate: these environments offered accessible participants, strong ecological validity for the researcher's own expertise, and rich opportunities to study planning bias in action. Nevertheless, future research should test whether the dynamics observed here, particularly the role of reflection and subjective realism, extend to other knowledge-intensive professions.

Sample sizes across the studies were modest, reflecting the resource constraints and exploratory aims of the project. While small samples limit statistical power and the ability to detect subtle effects, they were sufficient to capture qualitative insights into lived experiences and to trial novel interventions. The thesis prioritised depth of understanding over breadth of coverage, aligning with HCI traditions of exploratory, mixed-methods research. Larger-scale replication will be necessary to confirm the prevalence and robustness of the patterns identified here.

The intervention studies were relatively short-term, typically lasting two weeks. This duration allowed for manageable data collection and reduced participant burden, but it limits claims about long-term behaviour change. Planning habits are deeply ingrained, and it is possible that initial gains in self-efficacy or reduced procrastination may diminish over time. The choice of short interventions was pragmatic, enabling the testing of feasibility before committing to larger trials. A natural next step would be longitudinal studies examining whether reflective planning support can be sustained over months or years, and whether cumulative exposure leads to lasting shifts in bias or motivation.

The research combined self-report measures with some behavioural data, such as logged task durations. Self-report was necessary to capture subjective experiences of realism, motivation, and affect, which were central to the argument of this thesis. However, reliance on self-report introduces the risk of bias, and objective behavioural improvements in estimation accuracy remained elusive. This reflects both a limitation of the methods and a theoretical claim: that the value of planning cannot be captured solely through accuracy metrics. The emphasis on subjective outcomes was a considered choice, grounded in the empirical finding that participants reported benefits even when accuracy did not improve. Still, future work could expand the range of measures to include physiological, organisational, or long-term performance indicators, allowing for a richer triangulation of outcomes.

The interventions trialled were lightweight, focusing on simple strategies such as time tracking or feedback prompts. This narrow scope means that other potentially valuable approaches, such as collaborative planning, workplace-level interventions, or more sophisticated AI companions, were not explored. The choice to keep interventions minimal was intentional, reducing barriers to participation and allowing the research to isolate the effects of reflective prompts without confounds. However, this also means the findings represent only a slice of the possible design space. Future research should examine more ambitious interventions that address planning at organisational as well as individual levels.

Finally, it is important to acknowledge that the researcher's position within academia shaped both access to participants and interpretations of findings. Being embedded in the same cultural context as many participants offered ecological sensitivity but may also have introduced bias in framing and analysis. Reflexivity was practiced throughout, yet different contexts or researchers may have highlighted other dynamics.

In summary, the limitations of this thesis lie in its focus on academic contexts, modest and short-term samples, reliance on self-report, and narrow intervention scope. These choices were deliberate, reflecting the exploratory aims of the research and the methodological traditions of HCI. They enabled the development of new constructs and concepts, such as subjective realism and reflective planning, that warrant testing in broader, longer, and more diverse studies. By recognising these limitations, the thesis frames its contributions not as definitive answers but as foundations for further inquiry.

6.8 Future directions

6.8.1 Longer-term interventions

The interventions reported in this thesis were relatively short-term. Studies 5 and 6 showed that debiasing strategies such as task duration feedback could shift how participants evaluated their plans, increasing the proportion judged as realistic, and could prompt adaptations in planning style. However, the durability of these changes remains unknown. Planning is a habitual, long-horizon practice, and future work must investigate how reflective scaffolds unfold over months or years. Key questions include whether early benefits such as reduced procrastination and increased self-efficacy persist, whether cumulative improvements in estimation accuracy emerge with sustained use, and whether planning becomes more adaptive over time. Methodologically, longitudinal field deployments with mixed methods, combining trace data, diaries, and follow-up interviews, would provide rich evidence about both behavioural and experiential change.

6.8.2 Diverse work settings

This thesis focused on academia as a site of knowledge-intensive work. The choice was fruitful: academia shares features with many knowledge professions, including autonomy, competing demands, and blurred worklife boundaries. Nevertheless, academic work is distinctive, and future research should extend the study of optimistic planning bias into other professional settings. Clinical environments, for example, place greater emphasis on collaborative planning, while consultancy and corporate contexts may foreground shifting priorities and rapid rescheduling. Creative work, by contrast, often thrives on looser temporal structures [169]. Exploring reflective planning across these diverse settings would test the generalisability of the constructs developed here, such as subjective realism and reflective planning, and could refine them to account for contextual factors. This may involve comparative studies across sectors, or partnerships with organisations to embed reflective tools in live work systems.

6.8.3 Integration of reflection and values

A recurring finding of this thesis was that reflective engagement with plans was valuable even when objective accuracy did not improve. This points to the need for planning tools that do more than optimise efficiency: they should also help individuals align their daily commitments with what they find meaningful. Integrating reflective planning with value clarification would extend the design space from productivity support into wellbeing support. Questions for future work include: how can PTM systems make values visible without becoming burdensome? What forms of lightweight reflection help people connect micro-level planning with broader life priorities? And how can these features be evaluated in terms of meaningfulness as well as performance? Bridging literatures on meaningful work, wellbeing, and life crafting, future tools could support richer forms of engagement than efficiency alone.

6.8.4 AI-supported planning

The rise of generative AI raises new questions for planning support. While current productivity assistants tend to automate scheduling, our findings point instead to the possibility of AI systems that scaffold reflective engagement. Open research questions include how LLMs might be integrated without displacing user judgement, and how to design calibration strategies that prevent overreliance. Addressing these questions requires prototyping and participatory methods, with attention to trust and accountability.

6.8.5 Multi-level interventions

Finally, the findings highlight that planning is not only an individual practice but also an ecological one, shaped by organisational norms and systemic pressures. Participants' experiences of overcommitment and bias were tied to cultural expectations in academia around output, availability, and self-management. Interventions that focus solely on the individual risk obscuring these wider drivers. Future research should therefore investigate multi-level interventions that redistribute responsibility for sustainable planning. Examples include reflective check-ins integrated into collaboration platforms, dashboards that visualise workload distribution across teams, or institutional policies that protect time for planning and recovery. Such directions extend the scope of planning support beyond individual self-regulation to collective and structural change. They point to a research agenda that integrates individual, team, and policy perspectives to address the systemic drivers of unrealistic planning.

6.9 Conclusion

This thesis examined how planning technologies can support more accurate task planning. Across six studies, it showed that optimistic planning bias is a recurrent feature of academic work, shaping how participants structured their days, postponed tasks, and managed workloads that often exceeded capacity. Accuracy, defined as close correspondence between estimates and outcomes, proved difficult to achieve, and interventions focused narrowly on error correction often discouraged rather than supported participants. The thesis contributes in two main ways. First, it introduces reflective planning as a designable construct for understanding planning as an iterative process in which breakdowns and deferrals provide opportunities for adjustment. Second, it highlights the importance of what can be termed subjective realism: the perception that plans are realistic enough to sustain motivation and guide action. While exploratory, this construct draws attention to motivational and affective dimensions that are often overlooked in accounts of planning bias.These insights address gaps in both research and design. Existing task management tools typically emphasise capture, reminders, and throughput, while institutions often treat planning failures as individual shortcomings. The findings suggest that accuracy is better supported when technologies scaffold reflection, surface patterns of deferral, and legitimise recalibration. Implications follow at several levels. Universities can incorporate reflective planning into training; software developers can prioritise sustainability over acceleration; workplaces can recognise overcommitment as structural rather than individual; and policymakers can use planning data to inform reforms. In revisiting the central research question, the thesis concludes that planning technologies can support accuracy when they are designed to scaffold reflective engagement and promote a workable sense of realism. This perspective provides a foundation for developing systems that support productivity and wellbeing within contemporary knowledge work.

6.9.1 Acknowledgement of AI assistance

Portions of the writing process for this thesis were supported by the use of OpenAI's ChatGPT. The tool was used primarily for language editing, restructuring drafts, and generating alternative phrasings, while all conceptual contributions, empirical analyses, and interpretations remain my own. All final decisions about wording, framing, and argumentation were made by the author.

Appendix A

Corpus of PTM apps

Todoist	Trello	Asana	Microsoft To-Do	ClickUp
Sunsama	Habitica	Remember the milk	ToodleDo	TeuxDeux
TickTick	Lunatask	Superproductivity	Google Tasks	Quick To-Do
Any.do	Zenkit Projects	Zenkit To Do	nTask	Ayoa
Teamwork	I Done This	Nozbe	Nozbe Personal	Todo Cloud
Focuster	MyLifeOrganized	Apple reminders	Omnifocus	Serene
Suru	Sorted3	Zoho	Checklist	TaskPaper
Todo.txt	Hive	Jira	Wrike	Apollo
Evernote	Dynalist	Motion	Notion	Monday.com
Basecamp	Akiflow			

Table A.1: Corpus of PTM apps identified Study 4

Appendix B

Corpus of PTM functionalities

Functionality Group	Functionality	Definition
Hierarchies	Tasks	Tasks are units of activities referred to as "tasks", "to-dos", "activities", "items", "issues", etc. that users add by specifying a task title. Tasks are the building block of to-do lists. Tasks are usually added at the bottom or more rarely at the top of task lists, designated by a plus sign. Users can usually add task characteristics to tasks, such as due dates and priorities.

Functionality Group	Functionality	Definition
Hierarchies	Task lists	Lists are placeholders for multiple tasks. Lists are often referred to as "projects". Lists can also be referred to as "areas" or "contexts". Irrespective of how the app chooses the name them, each task list should contain a collection of tasks. In some rare cases, task lists are done on a board rather than on a list and this also qualifies for a task list. To earn a point for task lists, an app should be able to support users in having more than one task list.
Hierarchies	Checklist	Checklist are items within a task that can be ticked off. A checklist can also be a subtask depending if it meets the criteria for subtasks as well as for checklist.
Hierarchies	Subtasks	Subtasks allow users to break down tasks into smaller tasks. Subtasks can usually be found within a task dropdown (or just underneath a task) and can be assigned attributed such as due date or priority. Subtasks can be both a checklist (if they can be ticked off) and subtasks if they have attributes.

Functionality Group	Functionality	Definition
Hierarchies	Folders	Folders are placeholders for multiple task lists. They can be reffered to by different names, such as "folders", "portfolios" ,"projects" and other terms used for lists as well. What is important is whether multiple task lists can be grouped together into a folder.
Hierarchies	Infinite nesting	Infinite nesting means that subtasks of subtasks can be created infinitely. In other words, tasks can have subtasks and each subtask can have its own subtasks, too.
Hierarchies	Sections	Sections can be added within task lists and do not contain task attributes, such as due dates and priority. Sections are simply collapsable headers. They are usually added by a plus button and "add section" at the bottom of task lists. Sections help to clarify hierarchies within lists.
Timings	Due date	Due dates contain the date by which a task must be completed. They are usually found in the task dropdown menu and are usually referred to as "due dates". Users generally set due dates by picking one in a dropdown calendar window, or a scroll window.

Functionality Group	Functionality	Definition
Timings	Due hour	Due hour is same as due date but designates the hour rather than the date a task is due by. Due hour is sometimes difficult to locate and needs to be accessed by a button such as "add time" or "add due hour" at the bottom of the due date menu.
Timings	Start date	Shows the date a task must be started on. It is sometimes hidden in the same way a due hour is hidden out of the main box of due dates.
Timings	Start hour	Shows the hour a task must be started on. It is sometimes hidden in the same way a due hour is hidden out of the main box.
Timings	Recurrence	Recurrence means that a task can be set to recur at specific days times. It is sometimes located in the task dropdown menu and other times it is inside the due date (or just date or time) button within the same menu. If the app supports natural language processing, recurring tasks can be set by adding a phrase such as "every wednesday" in the task name.
Timings	Time tracking	Time tracking allows users to specify the time it takes them to work on a task and it is usually done by a timer for each task.

Functionality Group	Functionality	Definition
Timings	Time estimates	Time estimates allows users to try to predict the time it will take them to complete a task.
Timings	Dependencies	Dependencies are special relationships between tasks, such as one task "waiting on" another one, "blocking" one another, etc. and can be found within the task menu. They allow two tasks to be linked together by a temporal relationship that must be kept, for instance, one task must be finished for the next one to start (e.g. "book a hotel" should follow after "decide on a destination").
Properties	Completion	Completion allows users to mark a task as complete by ticking or crossing it off. If this can only be done by assigning a "completed" status or date, then it does not count as completion.
Properties	Tags and labels	A user-generated tag or in some cases called "label" or "context" that allows users to associate a task property of their choice with a task. It is sometimes done by a or @ sign, and other times it can be found in the task dropdown menu.
Properties	Priority	A default tag that denotes priority and has levels, from lower priority to higher priority.

Functionality Group	Functionality	Definition
Properties	Status	A default tag that denotes status, such as "in progress".
Properties	Flag, star or !	A default tag that can be assigned to tasks and consists of a flag, star and exclamation mark. It shows priority however it does not have levels (low or high).
Properties	Difficulty	A default tag that indicates difficulty.
Properties	Risk	A default tag that indicates risk.
Properties	Effort	A default tag that indicates effort.
Properties	Importance	A default tag that indicates importance.
Properties	Urgency	A default tag that indicates urgency.
Properties	Workload	A default tag that indicates workload.
Supporting Information	Notes and descriptions	Being able to leave a note or description for a task. Usually located underneath the task title in the task menu dropdown.
Supporting Information	Attachments (or photos)	Being able to attach documents or photos to a task. To ear a point here, an app should allow users to upload a file to their asks even if only photos.
Supporting Information	Location	Being able to associate a location to a task.

Functionality Group	Functionality	Definition
Views	List	A view of tasks that displays them all in one or more lists on the same page.
Views	Calendar	A view of tasks that displays them on a calendar.
Views	Table	A view of tasks that displays them in a table.
Views	Board	A view of a task list that shows its sections, groups or any other characteristic, such as progress, side by side and allows tasks to be moved across the sections by drag and drop.
Views	Gantt or timeline	A view of tasks that displays them on a timeline. A specific type of timeline is the Gannt chart.
Views	Dashboard	A dashboard is a high-level summary page that shows progress on tasks, graphs, summary statistics, etc. and it varies according to apps. It is normally referred to as "dashboard" but it could also have a different name, e.g. "projects".
Views	Folder	An option to view all tasks that are contained in a folder by clicking on the folder.
Views	Mindmap	A view of tasks that displays them in a mindmap.
Views	Map	A view of tasks that displays them on a map based on the location that users assign to them.

Functionality Group	Functionality	Definition
Browse	Search	Search allows tasks and other information within the app to be retrieved based on a name that they contain. It is usually done by a search bar.
Browse	Sort	Sorting tasks based on a criteria.
Browse	Filter	Filter refers to obtaining a view of only certan tasks based on a criteria, such as due date. It is done within a task list similarly to how filtering in word documents works.
Browse	Group by	Groups are similar to sections and allow tasks within a list to be grouped together by collapsable headers however those headers are based on a criteria, such as priority or deadline.
Browse	Multifilter	Filtering based on more than one criteria.
Browse	Multisort	Sorting tasks based on more than one criteria.
Completed Section	Completed tasks in a list	When marking a task as completed, it often goes in a separate section with all completed tasks within the task lists. Alternatively, the app may make it easy to see all compelted tasks within a list, for instance, by clicking on "show completed tasks". In other cases, a simple "sort" will work for clustering all compelted task in a list. As long as it is simple, it should count.

Functionality Group	Functionality	Definition
	Completed tasks in a folder	If app offers a folder view, it sometimes has a section of all completed tasks within the folder view.
Smart Lists	Today	A default list that displays automatically tasks due on the same day.
Smart Lists	All tasks	A default list that displays automatically tasks due on the next day.
Smart Lists	Calendar in a list	A default task list that displays automatically tasks sorted according to the date. It can be referred to as "Upcoming". It resembels a calendar however it looks more like a list of tasks than a traditional calendar view.
Smart Lists	Later	A default list that displays automatically all tasks due beyond a certain day (usually after today or tomorrow) or tasks that have a due date scheduled. It is referred to by different terms such as "Next", "Up Next", "Later", "Scheduled". This list does not have structure such as days of the week or dates.
Smart Lists	Inbox	A default list called "inbox" that contains all tasks that are not assigned a due date.
Smart Lists	Completed	A default list that displays automatically all completed tasks.

Functionality Group	Functionality	Definition
Smart Lists	Tomorrow	A default list that displays automatically tasks due on the next day.
Smart Lists	Overdue	A default list that displays automatically overdue tasks or tasks with missed reminders.
Smart Lists	Important	A mixed default list that displays automatically all tasks that can be considered important, such as overdue tasks and high priority tasks. It can be referred to as "Hotlist" or "Flagged", etc. as well as "Important".
Smart Lists	No due date	A default list that displays automatically all tasks that do not have a due date assigned. It can also be shows as "Unscheduled".
Smart Lists	Next 7 days / This week	A default list that displays automatically tasks due over the next 7 days or tasks due during the week.
Smart Lists	Assigned to me / Related to me / My work	A default list that displays automatically all tasks that are assigned to the user or the user is involved in. This list might be in "My work" section for more collaborative apps.
Smart Lists	Next 30 days	A default list that displays automatically tasks due in the next 30 days.
Smart Lists	Trash	A default list that displays automatically all deleted tasks.

Functionality Group	Functionality	Definition
Smart Lists	Won't do	A default list that shows all tasks that have been assigned the "won't do" label.
Smart Lists	Tags and labels	A default folder that displays automatically task lists based on all existing filters and labels.
Smart Lists	Assigned to others	A default list that displays automatically all tasks that are assigned to other team members.
Members	List sharing	Task lists that can be shared with others.
Members	Task assigning	Tasks can be assigned to others.
Members	Folder sharing	Folders that can be shared with others.
Members	Task sharing	Tasks within lists that can be shared with others.
Discussions	Comments	Comments can be left under each task. This is usually found in the task menu.
Discussions	Home (Feed)	A feed is a place for all posts and comments done by team members. Users can also sometimes post updates in there or other task information to share with others. It may be called "Home", etc.
Discussions	Mentions	Being able to mention a person within comments. Usually done by @ sign.
Discussions	Likes (Boosts)	Being able to like or give a boost for a task.

Functionality Group	Functionality	Definition
Discussions	Voting	Being able to vote for or give a boost on a task or comment on a task.
Collaborative Notifications	Notification section	A dropdown section that displays recent notifications and is called "notifications". It often uses a buzz icon.
Collaborative Notifications	Notification status	An indication in the notification section that shows if each notification has been seen/read by the user.
Collaborative Notifications	Follow tasks or lists	Optional notifications for certain tasks or lists. It is usually found in the task dropdown menu or list options. It can also be referred to as "watch" tasks or projects.
Collaborative Notifications	Do not disturb	Temporarily pausing notifications. It is usually done from notification section or notification settings. It can alse be referred to as "Mute".
Activity Log	General activity log	A historical log of all task operations. For instance, task created, comments, deletions, completions, etc. with a time stamp and the user who has made the changes.
Activity Log	Task activity log	A historical log of task history usually found in the task menu bar or just below comments.

Functionality Group	Functionality	Definition
AI support	Automatic task suggestions	Usually a functionality related to "today's" tasks that allows users to receive AI suggestions about which tasks to include in their day list. It could also be that users ask for a recommendation about what task to do over the next hour or so.
AI support	Scheduler	Scheduler allows users to auto-generate a daily schedule based on a list of tasks that users assign durations to.
AI support	Automatic scheduling shortcuts	When scheduling a due date, in some rare cases apps can give shortcut options (such as "tomorrow" or "today") based on AI argorithms that predicts which option is most likely to be suitable for a task.
Worklaod Management	Scheduled/Unscheduled time	This is a functionality which shows the amount of time that users have already scheduled and the respective amountof time left (unscheduled) for a given day or week.
Worklaod Management	Work-in-progress limit	Usually in settings, users can pick a limit for their tasks that can receive the "in-progress" status. For instance, no more than than 3 tasks can be started in the same time.

Functionality Group	Functionality	Definition
Customization Guidance	Suggestions for custom smart lists	When creating a new smart lists, apps can offer suggestions on what criteria to use.
Customization Guidance	Suggestions for automation (rules)	When creating rules to automate their workflow, apps may offer some templates to choose from to guide users towards the automations that would be useful.
Customization Guidance	Guided list set-up	When users create new task lists, sometimes the app can guide them to input key task properties. For instance, users may be shows questions such as "When is the final project delivery date" which will automaticelly input a final due date for the task list.
Customization Guidance	Productivity method quiz	Usually found on the app website, in some rare cases apps offer a quiz that recommends a productivity method (such as GTD or Pomodoro timers) to the users and it is explained how to implement it through the app.
Customization Guidance	Templates	Templates are ready-made task lists that have a specific purpose, such as a weekly review template or 1-on-1 meeting template. They are sometimes located in their own database within the app and other times users can pick one when creating a new list. Also, some apps offer them upon signing up.

Functionality Group	Functionality	Definition
Contact us	Email or message	When contacting the app customer support team, apps usully offer a message or email route in the "Contact us" section.
Contact us	Chat	Some apps offer a live chat function usually located in the lower right corner of the app window.
Contact us	Phone	Apps may offer phone customer support options. This is not the same as Demo Call. Phone support should be available more than once.
Contact us	Feature request	Feature requests are a special type of messages that the user sends to the app team. They allow users to request a feature that does not exist in the app yet and to sometimes vote on desired features that others have requested. It is also sometimes used as a place for the app team to provide information on which features they are working on and which are planned and at what poitns in the future. Sometimes, feature requests are done through the app wensite as a sub-topic in their forums.
Contact us	Report a bug	Reporting a bug is usually done through the app website and is called "report a bug".

Functionality Group	Functionality	Definition
Help Center	Blog	A blog is a website with information about the app and how to use it, and some tips and tricks about time maangement. It is found on the main app webpage and is referred to as "blog".
Help Center	Forum	Forum allows users to connect to one another and post and aswer questions about the app. To earn a point for a forum, an app should be the one that maintains it on their app platforms, for instance, on their website.
Help Center	FAQ	A FAQ section usually accessed through a "Help" button in the app.
Help Center	Screenshot tutorials	Apps may provide screenshots when instructing users how to use the app. Usually screenshot tutorials are found in the FAQ section.
Help Center	Video tutorials and webinars	Some apps upload videos and websinars on how to use the app and how to set it up, how to build a workflow, etc. This is usually found in FAQ and help centers.
Help Center	Integrations overview	Overview of native app integrations found in an "integrations" section.
Help Center	Integrations how-to guide	Apart from listing all possible integrations, some apps offer step-by-step guide on how to set up integrations.

Functionality Group	**Functionality**	**Definition**
Help Center	Onboarding walk-through or check-list	Upon signing up, apps often offer a tutorial or a checklist to help users learn how to use the app functionalities.
Help Center	University or certification course	Apps may give an in-depth education on how to use the app through a so called "university" and "academy" or certification course that can be found on their websites.
Help Center	Demo call	Demo calls are aimed at providing the user with a 1-on-1 explanation about how to set up and use the app over a call with an app member. They are usually 30 min in duration and are free for premium users.
Habits	Habit tracker	Habit traskers allow users to track how often (how many days) they have performed a habit of their choice.
Habits	Habit analytics	Habit analytics allow users to see data visualisations on trends in their habits.
Habits	Mood tracker	Mood trackers allow users to record and track their mood.

Functionality Group	Functionality	Definition
Rituals	Daily planning	Daily planning is a guided step-by-step questionnaire that users can use to make a plan for their day every day and prioritize their most important tasks. Sometimes, daily planning suggests tasks from other apps, e.g. calendar, that users may wish to include in their day.
Rituals	Daily shutdown	Daily shutdown is a guided step-by-step questionnaire that users can use to review their tasks for that day and detach from work.
Rituals	Weekly planning	Weekly planning asks users once a week to make a plan for the week ahead and prioritize their most important tasks.
Rituals	Weekly review	Weekly review is a functionality that supports users in finding the time to review and update their tasks lists or certain projects. Some apps allow users to pick a time for a weekly review for different tasks or projects, and the app would remind them to do the review. Other apps may directly guide the user toward reviewing all their task lists once a week, and may ask users to reflect on how their week went.
Completion rates	Completed tasks tracker	A count of completed tasks for a certain period, usually a week or a day.

Functionality Group	Functionality	Definition
Completion rates	Targets for completed tasks	When users can set target for completed tasks, for example, 5 each day.
Time Analytics	Time Analytics Metrics (task)	Metrics for tasks shows to the user various task data apart from completed task count that may be related to their performance, for instance, proportion of completed vs. in progress vs, incompleted tasks, proportion of high vs low priority tasks, etc. Metrics use numbers and do not use data visualisations (that is separate, see below).
Time Analytics	Time Analytics Metrics (time distribution)	Metrics for time distribution shows to the user various time tracking data that may be related to their performance, for instance, time spent across different projects, estimated vs actual time on tasks, etc. Metrics use numbers and do not use data visualisations (that is separate, see below).
Time Analytics	Time Analytics Data visualisations (tasks)	Data visualistions for tasks present the data used to task metrics but in graphs and charts.
Time Analytics	Time Analytics Data visualisations (time distribution)	Data visualistions for time distribution present the data used to time distribution metrics but in graphs and charts.

Functionality Group	Functionality	Definition
Goal-setting	Goals	Goals are always referred to as "goals" in apps and are separate from tasks. Users can set goals that go beyond single tasks and are usually longer term objectives.
Goal-setting	Sub-goals	Soma apps allow users to break down goals into sub-goals. Apps also may allow users to connect their tasks to goals or subgoals.
Distractions and Focus	Pomodoro session	Apps may have pomodoro timers that allow users to focus on work tasks for a certain amount of time, e.g. 25 min and to take a break, e.g. for 5 min.
Distractions and Focus	Task full screen	Some apps may offer to "zoom in" a task that the user wants to perform at that given moment. The app displays a full screen of that task without showing any information related to other tasks.
Distractions and Focus	Music and sounds	Apps may be able to play relaxing music or sounds, such as white noise, to help users focus on tasks.
Distractions and Focus	Simple timer	Apps may have a timer of different durations of choice (e.g. 10 min or 30 min) that can be used to count back time on tasks. Simple timers as opposed to pomodoro timers do not take breaks into account.

Functionality Group	Functionality	Definition
Distractions and Focus	Break reminder	Apps may have automatic break reminders.
Distractions and Focus	Distraction blocker	Apps may offer distraction blocking funtionailty. This may be done through a browser extension. The app blocks access to distracting websites, such as Youtube.
Distractions and Focus	Procrastination helper	Apps may help users use strategies to deal with procrastination while working.
Engagement	Gamification	Some apps turn planning into a game with points, characters or quests.
Engagement	Celebrations	Upon task completion some apps offer small encouragements such as celebration sounds, confetti or ballonons walking over the screen.
Drag and drop	Drag and drop to rearrange	Dragging and dropping tasks to rearrange their order in task lists.
Drag and drop	Drag and drop to calendar	Dragging and dropping tasks from a list to a calendar view, usually side by side.
Keyboard shortcuts	Shortcut keys	It is common for apps to offer keyboard shortcuts to navigate and interact with the app. Usually, these can be located in the setting.
Keyboard shortcuts	Command bar	Some apps may offer a command bar for quicker navigation and mouse free interactions.

Functionality Group	**Functionality**	**Definition**
Pinning	Pin a task	Apps may alow users to pin a task usually to the beginning of a task list.
Pinning	Favourite or pinned lists	Similarly, apps may alow users to pin a list (to make it a favourite) and see it in the main task list menu usually located on the left of the screen.
NLP	Smart date parsing	Date and time information users type in the name of the task are recognized as the due date.
NLP	Tag recognition	Tag information users type in the name of the task (usually following a or @ sign) is recognized as a tag. Also, priority pay be regognized with a p letter, e.g. p1 would be priority level 1 (low),
Reminders and Notifications	Reminders	Users can set a reminder in a time of their choice preceeding a task. At the time of the reminder, a notification will remind them that a task is due soon.
Reminders and Notifications	Location reminders	Some apps offer reminders based on a location, for instance, when at home or at the office. Some offer this only on the smartphone but that does not qualify for a point here.

Functionality Group	Functionality	Definition
Reminders and Notifications	Push notifications	Apps may notify users of updates by desktop pop-up notifications (push notifications). To enable them, users should normally allow them through the browser settings.
Reminders and Notifications	Email notifications	Some apps send notifications by email.
Reminders and Notifications	Daily digest	Daily digest is a summary email that the app sends the users which contains a highlight of daily task lists.
Reminders and Notifications	Interactivity in notifications	When a notification pops up on the screen, some apps allow users to either snooze them or mark them as seen, or some other interactivity.
Other quick functions	Add task name only	Adding a task title and pressing enter results in a new task rather than having to fill in other task details to save a task.
Other quick functions	Quick task summaries	When adding task properties such as due dates, some apps display them in a quick format underneath the task for a quick overview.
Other quick functions	Predefined options for due dates	When picking a due date, apps may offer predefined options, such as "today" or "tomorrow" or "next week" as opposed to dates and times only (e.g. 12 May 4PM).

Functionality Group	Functionality	Definition
Other quick functions	Predefined options for reminders	Similarly, apps sometimes provide options for setting a reminder time, such as "1 hour before the event" as opposed to a time and date (3PM on 12 May).
Other quick functions	Duplicate tasks and projects	Usually located in the task menu (three dots next to the task list or task), there might an option for copying (or duplicating) the task or task list.
Other quick functions	Move tasks	In the same menu, users may be able to move a task in a diffeernt task list.
Other quick functions	Batch edit	Batch edit is when more than one task can be editted at the same time. It is sometimes referred to as "bulk edit". Is it done either in the task menu or by holding the Ctrl key and selecting multiple tasks.
Other quick functions	One task in two lists	In some rare cases, apps may allow users to not only copy a task or move it between lists, but to have the same task appearing in two different lists.
Flexible Dates	Not-definite due dates	In some very rare cases, an app may offer a non fixed due date, for instance, before May 4.

Functionality Group	Functionality	Definition
Flexible Dates	Boomerang	Boomerang is a functionality that makes a task disappear for a certain amount of time during which users would not like it to be shown in their lists. After the time is over, it comes back (boomerangs).
Flexible Dates	Postpone	Once a task is due, some apps offer a shortcut button to postpone it to a future date.
Time Settings	Time zone	In the app settings, some apps allow users to change the time zone.
Time Settings	Week starts on	Similarly, some apps allow users to pick a first day of the week (e.g. Monday or Sunday).
Time Settings	Time format	Apps may allow the changing of the time format.
Time Settings	Date format	Apps may allow the changing of the date format.
Appearance	Dark mode	Dark mode switches the appearance to dark theme colours. It is usually found in the settings menu.
Appearance	Themes	Themes go beyond dark and light mode. They usually can also be found in the settings menu.
Appearance	Tag or filter colours	Some apps allow users to set colours for different tags, filters and categories.

Functionality Group	Functionality	Definition
Appearance	Emojis in titles	Some tasks have an emoji menu when creating a task.
Appearance	Background or task cover image	Some apps allow users to pick an image from a database or to upload their own image as a backgorund or as a cover to specific tasks.
Custom Properties	Change priority	Priority can sometimes be modified beyond the app levels of priority (usually three levels of low, medium and high). For instace, users can add extra levels of priorities and pick a name for those new levels, such as "extremely high".
Custom Properties	Change status	Similarly to priority, status can also be sometimes changed according to user preferences. For instace, an app may offer three simple status such as "not started", "in progress" and "completed" and users may be able to add additional ones of their choice.
Custom Views	Custom smart list or saved filters	Some apps allow users to create their own smart lists beyond the app ones. For instace, tasks that have certain tags or labels and associated priority or due dates. This is sometimes done by creating a smart list and other times is referred to as "save filters" or "save filter view". Both of these produce the same results.

Functionality Group	Functionality	Definition
Custom Views	List-based views	List-based views is a functionality that preserves the view of choice for each task list. For instance, if a user chages the task view to calendar, when the user exits and open again this list, the view is kept the same.
Custom Views	Custom dashboard	Some apps allow the customization of the home view. Users can pick which widgets to be displayed.
Custom Settings	Rearrange menus	This funtion allows users to rearrange the task menu, for instance, to switch the places of due date and priority boxes within the dropbdown.
Custom Settings	Custom shortcut keys	Some apps allow users to change the app shortcut keys to ones of their choice. This is done in settings.
Custom Settings	Custom events for notifications	In the notification settings, some apps allow users to pick events for which to receive notifications from a list.
Custom Settings	Default list	Some apps have default lists. This means that everytime the user adds a task when no list is specified, it will be added to a list of their choice referred to as "default list". This is done in settings.

Functionality Group	Functionality	Definition
Custom Settings	Home default	Some apps alllow users to select a default home page. That is, when users open the app, they would see that particular page, e.g. Inbox or a project.
Custom Settings	Task defaults	Similarly to default lists, some apps allow users to pick default task properties for new tasks. For instace, when users create a task, the app will automatically assign due date or priority, etc.
Custom functionalities	Directory of third-party add-ons	Some apps allow third parties (or users) to create add-ons. These third-party apps are presented in a directory. It is usually accessed in "Apps" section sometimes in Setting and other times on the main app homepage.
Custom functionalities	Automation (rules)	Automation or often referred to as "rules" is when users can automate their workflows by creating rules with triggers and events.
Custom functionalities	In-apps	In-apps allow users to select which funtionalities to include in the app. For example, repeating a task or a workload indicator. In some cases, in-apps are referred to as "in-apps" but in others, these are simply funtionalities that users can choose to turn on or off from settings.

Functionality Group	Functionality	Definition
Custom data visualisation	Custom charts and graphs	Some apps not only display data about time and tasks but also allow users to choose the parameters and charts of these visualisations.
Email-In	Email In	Email-in is a functionality that allows users to send emails as tasks. It is done by a special app email address that can be found in settings.
Native Integrations	Outlook	Apps may offer a native Outlook integration.
Native Integrations	Gmail	Apps may offer a native Gmail integration.
Native Integrations	Google calendar	Apps may offer a native Google Calendar integration.
Native Integrations	Slack	Apps may offer a native Slack integration.
Native Integrations	Teams	Apps may offer a native Teams integration.
Custom Integrations	Third party automation	Apps may offer custom integration through third parties, such as Zapier.
Custom Integrations	Two-way sync	Apps may offer custom two-way integration through third parties, such as Unito or Pleexy.
Add-ons and add-ins	Browser extension	Apps may offer browser extensions.
Add-ons and add-ins	Gmail add-on	Apps may offer a Gmail add-on.

Functionality Group	Functionality	Definition
Add-ons and add-ins	Outlook add-in	Apps may offer an Outlook add-in.
Add-ons and add-ins	Word add-in	Apps may offer a MS Word add-in.
Non-digital	Print	Apps may offer a "Print" button in the task menu.

Table B.1: Codebook with all functionalities identified in Study 4

Bibliography

[1] Victoria Bellotti et al. “What a to-do: studies of task management towards the design of a personal task list manager”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2004, pp. 735–742.

[2] Amirrudin Kamsin, Ann Blandford, and Anna L Cox. “Personal task management: my tools fall apart when I’m very busy!” In: *CHI’12 Extended Abstracts on Human Factors in Computing Systems*. 2012, pp. 1369–1374.

[3] Brigitte JC Claessens et al. “Things to do today...: A daily diary study on task completion at work”. In: *Applied Psychology* 59.2 (2010), pp. 273–295.

[4] Brigitte JC Claessens et al. “A review of the time management literature”. In: *Personnel review* 36.2 (2007), pp. 255–276.

[5] Christine J Syrek et al. “Zeigarniks sleepless nights: How unfinished tasks at the end of the week impair employee sleep on the weekend through rumination.” In: *Journal of Occupational Health Psychology* 22.2 (2017), p. 225.

[6] Donghan Hu et al. “Unpacking Task Management Tools, Values, and Worker Dynamics”. In: *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. 2024, pp. 1–16.

[7] Roger Buehler and Dale Griffin. “The planning fallacy”. In: *The psychology of planning in organizations: Research and applications* 31 (2015).

[8] Daniel Bennett and Elisa D Mekler. “Beyond intrinsic motivation: The role of autonomous motivation in user experience”. In: *ACM Transactions on Computer-Human Interaction* 31.5 (2024), pp. 1–41.

[9] Ann E Blandford and Thomas RG Green. “Group and individual time management tools: what you get is not what you need”. In: *Personal and Ubiquitous Computing* 5 (2001), pp. 213–230.

[10] Eva Geurts et al. *HCI and worker well-being*. 2024.

[11] Helen Nicholls et al. “The impact of working in academia on researchers mental health and well-being: A systematic review and qualitative meta-synthesis”. In: *PloS one* 17.5 (2022), e0268890.

[12] Gloria Mark. *Multitasking in the digital age*. Springer Nature, 2022.

[13] Judith Borghouts, Duncan P Brumby, and Anna L Cox. “TimeToFocus: feedback on interruption durations discourages distractions and shortens interruptions”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 27.5 (2020), pp. 1–31.

[14] Alan Bryman. *Social research methods*. Oxford university press, 2016.

[15] William M Newman. “Busy days: exposing temporal metrics, problems and elasticities through diary studies”. In: *CHI 2004 Workshop on Temporal Issues in Work. Citeseer*. Citeseer. 2004.

[16] Matthew J Page et al. “The PRISMA 2020 statement: an updated guideline for reporting systematic reviews”. In: *Bmj* 372 (2021).

[17] Elahi Hossain et al. “Mapping the landscape of functionality reviews in HCI: trends, methods, and contributions”. In: *Interacting with Computers* (2025), iwaf027.

[18] Alan Johnston and Eleanor Davies. “Academics psychological contract: applying discretionary effort”. In: *Journal of Further and Higher Education* (2025), pp. 1–15.

[19] Zahra Shakeri Hossein Abad et al. “Task interruption in software development projects: What makes some interruptions more disruptive than others?” In: *Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018*. 2018, pp. 122–132.

[20] Anita L Tucker and Steven J Spear. “Operational failures and interruptions in hospital nursing”. In: *Health services research* 41.3p1 (2006), pp. 643–662.

[21] Bruce Alberts et al. “Rescuing US biomedical research from its systemic flaws”. In: *Proceedings of the National Academy of Sciences* 111.16 (2014), pp. 5773–5777.

[22] Tait D Shanafelt et al. “Burnout and satisfaction with work-life balance among US physicians relative to the general US population”. In: *Archives of internal medicine* 172.18 (2012), pp. 1377–1385.

[23] George A. Miller, Eugene Galanter, and Karl H. Pribram. *Plans and the Structure of Behavior*. New York: Holt, Rinehart and Winston, 1960.

[24] Elisabeth Parés-Pujolràs et al. “Evidence accumulation under uncertainty-a neural marker of emerging choice and urgency”. In: *NeuroImage* 232 (2021), p. 117863.

[25] Edwin A Locke. “Motivation by goal setting”. In: *Handbook of organizational behavior* (2000), pp. 43–56.

[26] John R Lund and Jason Wiese. “Less is more: exploring support for time management planning”. In: *Designing Interactive Systems Conference 2021*. 2021, pp. 392–405.

[27] André N Meyer et al. “Enabling good work habits in software developers through reflective goal-setting”. In: *IEEE Transactions on Software Engineering* 47.9 (2019), pp. 1872–1885.

[28] Hayley Guillou et al. “Is your time well spent? reflecting on knowledge work more holistically”. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–9.

[29] Caroline Knight, Malcolm Patterson, and Jeremy Dawson. “Work engagement interventions can be effective: a systematic review”. In: *European Journal of Work and Organizational Psychology* 28.3 (2019), pp. 348–372.

[30] Young-Ho Kim et al. “TimeAware: Leveraging framing effects to enhance personal productivity”. In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2016, pp. 272–283.

[31] Peter M Gollwitzer and Paschal Sheeran. “Implementation intentions and goal achievement: A meta-analysis of effects and processes”. In: *Advances in experimental social psychology* 38 (2006), pp. 69–119.

[32] Yasaman S Sefidgar et al. “Improving Work-Nonwork Balance with Data-Driven Implementation Intention and Mental Contrasting”. In: *Proceedings of the ACM on human-computer interaction* 8.CSCW1 (2024), pp. 1–29.

[33] Albert Bandura. “Social cognitive theory of self-regulation”. In: *Organizational behavior and human decision processes* 50.2 (1991), pp. 248–287.

[34] Dale H Schunk and Barry J Zimmerman. *Motivation and self-regulated learning: Theory, research, and applications*. Routledge, 2012.

[35] Todd Rogers et al. “Beyond good intentions: Prompting people to make plans improves follow-through on important tasks”. In: *Behavioral Science & Policy* 1.2 (2015), pp. 33–41.

[36] James T Austin and Jeffrey B Vancouver. “Goal constructs in psychology: Structure, process, and content.” In: *Psychological bulletin* 120.3 (1996), p. 338.

[37] EJ Masicampo and Roy F Baumeister. “Consider it done! Plan making can eliminate the cognitive effects of unfulfilled goals.” In: *Journal of personality and social psychology* 101.4 (2011), p. 667.

[38] Sophie Leroy. “Why is it so hard to do my work? The challenge of attention residue when switching between work tasks”. In: *Organizational Behavior and Human Decision Processes* 109.2 (2009), pp. 168–181.

[39] Tina Ekhtiar et al. “Goals for goal setting: a scoping review on personal informatics”. In: *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 2023, pp. 2625–2641.

[40] Elena Agapie et al. “A longitudinal goal setting model for addressing complex personal problems in mental health”. In: *Proceedings of the ACM on Human-Computer Interaction* 6.CSCW2 (2022), pp. 1–28.

[41] John Rooksby et al. “Personal tracking as lived informatics”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2014, pp. 1163–1172.

[42] Michelle Y Tytherleigh* et al. “Occupational stress in UK higher education institutions: A comparative study of all staff categories”. In: *Higher education research & development* 24.1 (2005), pp. 41–61.

[43] Nicolene Barkhuizen and Sebastiaan Rothmann. “Occupational stress of academic staff in South African higher education institutions”. In: *South African journal of psychology* 38.2 (2008), pp. 321–336.

[44] Caroline Biron, Jean-Pierre Brun, and Hans Ivers. “Extent and sources of occupational stress in university staff”. In: *Work* 30.4 (2008), pp. 511–522.

[45] Peter James Bentley and Svein Kyvik. "Academic work from a comparative perspective: A survey of faculty working time across 13 countries". In: *Higher Education* 63 (2012), pp. 529–547.

[46] Gail Kinman. "Doing more with less? Work and wellbeing in academics". In: *Somatechnics* 4.2 (2014), pp. 219–235.

[47] Helen R Winefield, Carolyn Boyd, and Anthony H Winefield. "Work-family conflict and well-being in university employees". In: *The Journal of psychology* 148.6 (2014), pp. 683–697.

[48] Rita Fontinha, Darren Van Laar, and Simon Easton. "Quality of working life of academics and researchers in the UK: the roles of contract type, tenure and university ranking". In: *Studies in Higher Education* 43.4 (2018), pp. 786–806.

[49] Gail Kinman. "Pressure points: A review of research on stressors and strains in UK academics". In: *Educational psychology* 21.4 (2001), pp. 473–492.

[50] Stewart Collins and Beth Parry-Jones. "Stress: the perceptions of social work lectures in Britain". In: *British Journal of social work* 30.6 (2000), pp. 769–794.

[51] Victoria Hogan et al. "An examination of gender differences in the impact of individual and organisational factors on work hours, work-life conflict and psychological strain in academics". In: *The Irish Journal of Psychology* 35.2-3 (2014), pp. 133–150.

[52] Mary Henkel. "Academic identity and autonomy in a changing policy environment". In: *Higher education* 49 (2005), pp. 155–176.

[53] Abraham Mhaidli and Kat Roemmich. "Overworking in HCI: A reflection on why we are burned out, stressed, and out of control; and what we can do about it". In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–10.

[54] Mona Haraty, Joanna McGrenere, and Charlotte Tang. "How and why personal task management behaviors change over time". In: *Proceedings of the 41st Graphics Interface Conference*. 2015, pp. 147–154.

[55] Mona Haraty and Joanna McGrenere. "Designing for advanced personalization in personal task management". In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 2016, pp. 239–250.

[56] Ann Blandford, Dominic Furniss, and Stephann Makri. *Qualitative HCI research: Going behind the scenes*. Morgan & Claypool Publishers, 2016.

[57] Amos Tversky and Daniel Kahneman. “Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.” In: *science* 185.4157 (1974), pp. 1124–1131.

[58] Tim Reeves and Robert S Lockhart. “Distributional versus singular approaches to probability and errors in probabilistic reasoning.” In: *Journal of Experimental Psychology: General* 122.2 (1993), p. 207.

[59] Michael M Roy, Nicholas JS Christenfeld, and Craig RM McKenzie. “Underestimating the duration of future events: Memory incorrectly used or memory bias?” In: *Psychological bulletin* 131.5 (2005), p. 738.

[60] Torleif Halkjelsvik and Magne Jørgensen. “From origami to software development: A review of studies on judgment-based predictions of performance time.” In: *Psychological bulletin* 138.2 (2012), p. 238.

[61] Roger Buehler, Dale Griffin, and Michael Ross. “Exploring the" planning fallacy": Why people underestimate their task completion times.” In: *Journal of personality and social psychology* 67.3 (1994), p. 366.

[62] Magne Jorgensen and Stein Grimstad. “Software development estimation biases: The role of interdependence”. In: *IEEE Transactions on Software Engineering* 38.3 (2012), pp. 677–693.

[63] Jakob E Bardram and Thomas Riisgaard Hansen. “Why the plan doesn’t hold: a study of situated planning, articulation and coordination work in a surgical ward”. In: *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. 2010, pp. 331–340.

[64] Stephen R Covey, A Roger Merrill, and Rebecca R Merrill. *First things first*. Simon and Schuster, 1995.

[65] NS Jyothi and A Parkavi. “A study on task management system”. In: *2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS)*. IEEE. 2016, pp. 1–6.

[66] David Allen. *Getting things done: The art of stress-free productivity*. Penguin, 2015.

[67] Edwin A Locke and Gary P Latham. “Goal setting theory: The current state”. In: *New developments in goal setting and task performance*. Routledge, 2013, pp. 623–630.

[68] Terry Connolly and Doug Dean. “Decomposed versus holistic estimates of effort required for software writing tasks”. In: *Management Science* 43.7 (1997), pp. 1029–1045.

[69] Justin Kruger and Matt Evans. “If you don’t want to be late, enumerate: Unpacking reduces the planning fallacy”. In: *Journal of Experimental Social Psychology* 40.5 (2004), pp. 586–598.

[70] Ian R Newby-Clark et al. “People focus on optimistic scenarios and disregard pessimistic scenarios while predicting task completion times.” In: *Journal of Experimental Psychology: Applied* 6.3 (2000), p. 171.

[71] Simon Tobin and Simon Grondin. “Prior task experience affects temporal prediction and estimation”. In: *Frontiers in psychology* 6 (2015), p. 916.

[72] Dale Griffin and Roger Buehler. “Frequency, probability, and prediction: Easy solutions to cognitive illusions?” In: *Cognitive psychology* 38.1 (1999), pp. 48–78.

[73] Morgane Koval and Yvonne Jansen. “Do You See What You Mean? Using Predictive Visualizations to Reduce Optimism in Duration Estimates”. In: *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 2022, pp. 1–19.

[74] Yan Zhou et al. “Estimating usage can reduce the stress of social networking”. In: (2013).

[75] Anna L Cox, Jon Bird, and Rowanne Fleck. “Digital epiphanies: how self-knowledge can change habits and our attitudes towards them”. In: (2013).

[76] Yoana Ahmetoglu, Duncan P Brumby, and Anna L Cox. “Time Estimation Bias in Knowledge Work: Tasks With Fewer Time Constraints Are More Error-Prone”. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–8.

[77] Yoana Ahmetoglu, Duncan P Brumby, and Anna L Cox. “To plan or not to plan? A mixed-methods diary study examining when, how and why knowledge work planning is inaccurate”. In: *Proceedings of the ACM on Human-Computer Interaction* 4.CSCW3 (2021), pp. 1–20.

[78] Yoana Ahmetoglu, Duncan P Brumby, and Anna L Cox. “Disengaged from planning during the lockdown? an interview study in an academic setting”. In: *IEEE Pervasive Computing* 20.4 (2021), pp. 18–25.

[79] Kjetil Moløkken and Magne Jørgensen. “A review of surveys on software effort estimation.(2003)”. In: (2003).

[80] Lucille Alice Suchman. *Plans and situated actions: The problem of human-machine communication*. Cambridge university press, 1987.

[81] Roger Buehler, Dale Griffin, and Johanna Peetz. “The planning fallacy: Cognitive, motivational, and social origins”. In: *Advances in experimental social psychology*. Vol. 43. Elsevier, 2010, pp. 1–62.

[82] Ilaria Redaelli and Antonella Carassa. “New perspectives on plans: Studying planning as an instance of instructed action”. In: *Computer Supported Cooperative Work (CSCW)* 27 (2018), pp. 107–148.

[83] Cal Newport. *Deep work: Rules for focused success in a distracted world*. Hachette UK, 2016.

[84] Joseph D Blackburn, Gary D Scudder, and Luk N Van Wassenhove. “Improving speed and productivity of software development: a global survey of software developers”. In: *IEEE transactions on software engineering* 22.12 (1996), pp. 875–885.

[85] Adam Bradley et al. “How to manage your inbox: Is a once a day strategy best?” In: BCS Learning and Development Limited. 2013.

[86] Gloria Mark et al. “Email duration, batching and self-interruption: Patterns of email use on productivity and stress”. In: *Proceedings of the 2016 CHI conference on human factors in computing systems*. 2016, pp. 1717–1728.

[87] Margery Eldridge and William Newman. “Agenda benders: modelling the disruptions caused by technology failures in the workplace”. In: *Conference Companion on Human Factors in Computing Systems*. 1996, pp. 219–220.

[88] Victor M González and Gloria Mark. “" Constant, constant, multi-tasking craziness" managing multiple working spheres”. In: *Proceedings of the SIGCHI conference on Human factors in computing systems*. 2004, pp. 113–120.

[89] Jing Jin and Laura A Dabbish. “Self-interruption on the computer: a typology of discretionary task interleaving”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2009, pp. 1799–1808.

[90] Darryl K Forsyth and Christopher DB Burt. “Allocating time to future tasks: The effect of task segmentation on planning fallacy bias”. In: *Memory & cognition* 36 (2008), pp. 791–798.

[91] Johanna Peetz, Roger Buehler, and Anne Wilson. “Planning for the near and distant future: How does temporal distance affect task completion predictions?” In: *Journal of Experimental Social Psychology* 46.5 (2010), pp. 709–720.

[92] Lawrence J Sanna et al. “The hourglass is half full or half empty: temporal framing and the group planning fallacy.” In: *Group Dynamics: Theory, Research, and Practice* 9.3 (2005), p. 173.

[93] Sander Koole and Mascha van’t Spijker. “Overcoming the planning fallacy through willpower: Effects of implementation intentions on actual and predicted task-completion times”. In: *European Journal of Social Psychology* 30.6 (2000), pp. 873–888.

[94] Therese Hoff Macan. “Time management: Test of a process model.” In: *Journal of applied psychology* 79.3 (1994), p. 381.

[95] Anna Rudnicka et al. “Eworklife: developing effective strategies for remote working during the COVID-19 pandemic”. In: Aug. 2020.

[96] Jenna Butler and Sonia Jaffe. “Challenges and Gratitude: A Diary Study of Software Engineers Working From Home During Covid-19 Pandemic”. In: *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. 2021, pp. 362–363. DOI: 10.1109/ICSE-SEIP52600.2021.00047.

[97] Anna L. Cox et al. “Prioritizing Unread Emails: An Experiment Showing that People Send Urgent Responses Before Important or Short Ones”. In: *Human-Computer Interaction* (2020), pp. 1–24. DOI: 10.1080/07370024.2020.1846388. URL: https://doi.org/10.1080/07370024.2020.1846388.

[98] Rowan Fleck, Maria E. Cecchinato, Anna L. Cox, et al. “Life-swap: How Discussions Around Personal Data Can Motivate Desire for Change”. In: *Personal and Ubiquitous Computing* 24 (2020), pp. 669–681. DOI: 10.1007/s00779-020-01372-9. URL: https://doi.org/10.1007/s00779-020-01372-9.

[99] Alex C Williams et al. "Supporting workplace detachment and reattachment with conversational intelligence". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–13.

[100] Yoana Ahmetoglu, Duncan Brumby, and Anna Cox. "Bridging the Gap Between Time Management Research and Task Management App Design: A Study on the Integration of Planning Fallacy Mitigation Strategies". In: *Proceedings of the 3rd Annual Meeting of the Symposium on Human-Computer Interaction for Work*. CHIWORK '24. Newcastle upon Tyne, United Kingdom: Association for Computing Machinery, 2024. ISBN: 9798400710179. DOI: 10.1145/3663384.3663404. URL: https://doi.org/10.1145/3663384.3663404.

[101] Beata Smela et al. "Rapid literature review: definition and methodology". In: *Journal of Market Access & Health Policy* 11.1 (2023), p. 2241234.

[102] Timothy A Pychyl, Richard W Morin, and Brian R Salmon. "Procrastination and the planning fallacy: An examination of the study habits of university students." In: *Journal of Social Behavior & Personality* 15.5 (2000).

[103] Kevin Thomas, Simon Handley, and Stephen Newstead. "The effects of prior experience on estimating the duration of simple tasks." In: *Current Psychology of Cognition* 22.2 (2004), pp. 83–100.

[104] Kevin E Thomas, Simon J Handley, and Stephen E Newstead. "The role of prior task experience in temporal misestimation". In: *Quarterly Journal of Experimental Psychology* 60.2 (2007), pp. 230–240.

[105] Cornelius J König et al. "The effects of previous misestimation of task duration on estimating future task duration". In: *Current Psychology* 34 (2015), pp. 1–13.

[106] Anna C Brady, Christopher A Wolters, and Shirley L Yu. "Self-regulation of time: The importance of time estimation accuracy". In: *Frontiers in Psychology* 13 (2022), p. 925812.

[107] Michael M Roy, Scott T Mitten, and Nicholas JS Christenfeld. "Correcting memory improves accuracy of predicted task duration." In: *Journal of Experimental Psychology: Applied* 14.3 (2008), p. 266.

[108] Tanja M Gruschke and Magne Jørgensen. “The role of outcome feedback in improving the uncertainty assessment of software development effort estimates”. In: *ACM Transactions on Software Engineering and Methodology (TOSEM)* 17.4 (2008), pp. 1–35.

[109] Magne Jørgensen and Tanja M Gruschke. “The impact of lessons-learned sessions on effort estimation and uncertainty assessments”. In: *IEEE Transactions on Software Engineering* 35.3 (2009), pp. 368–383.

[110] Ofira Shmueli, Nava Pliskin, and Lior Fink. “Can the outside-view approach improve planning decisions in software development projects?” In: *Information Systems Journal* 26.4 (2016), pp. 395–418.

[111] Rebecca Friesdorf. “Using Our Past to Predict Our Future: Applying Reference Class Forecasting to Debias Individual Project Completion Predictions”. In: (2020).

[112] Bent Flyvbjerg. “Curbing optimism bias and strategic misrepresentation in planning: Reference class forecasting in practice”. In: *European planning studies* 16.1 (2008), pp. 3–21.

[113] Bent Flyvbjerg, Massimo Garbuio, and Dan Lovallo. “Delusion and deception in large infrastructure projects: two models for explaining and preventing executive disaster”. In: *California management review* 51.2 (2009), pp. 170–194.

[114] Ryen W White and Ahmed Hassan Awadallah. “Task duration estimation”. In: *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. 2019, pp. 636–644.

[115] Constantinos Hadjichristidis, Barbara Summers, and Kevin Thomas. “Unpacking estimates of task duration: The role of typicality and temporality”. In: *Journal of Experimental Social Psychology* 51 (2014), pp. 45–50.

[116] Rafay A Siddiqui, Frank May, and Ashwani Monga. “Reversals of task duration estimates: Thinking how rather than why shrinks duration estimates for simple tasks, but elongates estimates for complex tasks”. In: *Journal of Experimental Social Psychology* 50 (2014), pp. 184–189.

[117] Jason Wiese, John R Lund, and Kazi Sinthia Kabir. “Adding Domain-Specific Features to a Text-Editor to Support Diverse, Real-World Approaches to Time Management Planning”. In:

Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems. 2023, pp. 1–13.

[118] Roger Buehler and Dale Griffin. “Planning, personality, and prediction: The role of future focus in optimistic time predictions”. In: *Organizational Behavior and Human Decision Processes* 92.1-2 (2003), pp. 80–90.

[119] Stephanie J Byram. “Cognitive and motivational factors influencing time prediction.” In: *Journal of Experimental Psychology: Applied* 3.3 (1997), p. 216.

[120] Anat Bracha and Donald J Brown. “Affective decision making: A theory of optimism bias”. In: *Games and Economic Behavior* 75.1 (2012), pp. 67–80.

[121] Rebecca A Henry and Janet A Sniezek. “Situational factors affecting judgments of future performance”. In: *Organizational Behavior and Human Decision Processes* 54.1 (1993), pp. 104–132.

[122] Mario Weick and Ana Guinote. “How long will it take? Power biases time predictions”. In: *Journal of Experimental Social Psychology* 46.4 (2010), pp. 595–604.

[123] Erik Løhre and Karl Halvor Teigen. “How fast can you (possibly) do it, or how long will it (certainly) take? Communicating uncertain estimates of performance time”. In: *Acta Psychologica* 148 (2014), pp. 63–73.

[124] Roger Buehler et al. “Perspectives on prediction: Does third-person imagery improve task completion estimates?” In: *Organizational Behavior and Human Decision Processes* 117.1 (2012), pp. 138–149.

[125] Michael M Roy, Nicholas JS Christenfeld, and Meghan Jones. “Actors, observers, and the estimation of task duration”. In: *Quarterly Journal of Experimental Psychology* 66.1 (2013), pp. 121–137.

[126] Katarzyna Stawarz, Anna L Cox, and Ann Blandford. “Beyond self-tracking and reminders: designing smartphone apps that support habit formation”. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2015, pp. 2653–2662.

[127] Katarzyna Stawarz, Anna L Cox, and Ann Blandford. “Don’t forget your pill! Designing effective medication reminder apps that support users’ daily routines”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014, pp. 2269–2278.

[128] Ulrik Lyngs et al. “Self-control in cyberspace: Applying dual systems theory to a review of digital self-control tools”. In: *proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–18.

[129] Nora Ptakauskaite, Anna L Cox, and Nadia Berthouze. “Knowing what you’re doing or knowing what to do: how stress management apps support reflection and behaviour change”. In: *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–6.

[130] Alaa AlSlaity et al. “Mobile applications for health and wellness: a systematic review”. In: *Proceedings of the ACM on Human-Computer Interaction* 6.EICS (2022), pp. 1–29.

[131] Dilisha Patel, Ann Blandford, and Judith Stephenson. “So you’re planning a baby? a review of preconception care apps”. In: *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. 2018, pp. 63–70.

[132] Cameron Lister et al. “Just a fad? Gamification in health and fitness apps”. In: *JMIR serious games* 2.2 (2014), e3413.

[133] Rebecca Gulotta et al. “Fostering engagement with personal informatics systems”. In: *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 2016, pp. 286–300.

[134] N Ptakauskaite et al. “Personal informatics tools benefit from combining automatic and manual data capture in the long-term”. In: Association for Computing Machinery (ACM). 2018.

[135] Emily IM Collins et al. “Social networking use and RescueTime: the issue of engagement”. In: *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. 2014, pp. 687–690.

[136] Larry D Burton. *Can a Tomato Increase Your Productivity?* 2016.

[137] Miriam Jaffe et al. “Collaboration and potential space: creative play in the writing alliance”. In: *Teaching in Higher Education* (2021), pp. 1–16.

[138] Kaitlin Woolley and Ayelet Fishbach. “Its about time: Earlier rewards increase intrinsic motivation.” In: *Journal of personality and social psychology* 114.6 (2018), p. 877.

[139] Everlyne Kimani et al. “A conversational agent in support of productivity and wellbeing at work”. In: *2019 8th international conference on affective computing and intelligent interaction (ACII)*. IEEE. 2019, pp. 1–7.

[140] Yoana Ahmetoglu et al. “" I Work Much Better by Doing Less": How Task Duration Feedback Affects Optimistic Planning Bias in Academic Work”. In: *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work*. 2025, pp. 1–17.

[141] Ronald D Yockey. “Validation of the short form of the academic procrastination scale”. In: *Psychological reports* 118.1 (2016), pp. 171–179.

[142] Matthias Romppel et al. “A short form of the General Self-Efficacy Scale (GSE-6): Development, psychometric properties and validity in an intercultural non-clinical sample and a sample of patients at risk for heart failure”. In: *GMS Psycho-Social-Medicine* 10 (2013).

[143] Victoria Clarke and Virginia Braun. “Thematic analysis”. In: *The journal of positive psychology* 12.3 (2017), pp. 297–298.

[144] Virginia Braun and Victoria Clarke. “Reflecting on reflexive thematic analysis”. In: *Qualitative research in sport, exercise and health* 11.4 (2019), pp. 589–597.

[145] Falko F Sniehotta et al. “Action planning and coping planning for long-term lifestyle change: theory and assessment”. In: *European journal of social psychology* 35.4 (2005), pp. 565–576.

[146] Dario D Salvucci and Niels A Taatgen. “Threaded cognition: an integrated theory of concurrent multitasking.” In: *Psychological review* 115.1 (2008), p. 101.

[147] Martin T Orne. “Demand characteristics and the concept of quasi-controls”. In: *Artifacts in behavioral research: Robert Rosenthal and Ralph L. Rosnows classic books* 110 (2009), pp. 110–137.

[148] Young-Ho Kim et al. “Understanding personal productivity: How knowledge workers define, evaluate, and reflect on their productivity”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–12.

[149] Elisabeth T Kersten-van Dijk et al. “Personal informatics, self-insight, and behavior change: A critical review of current literature”. In: *Human–Computer Interaction* 32.5-6 (2017), pp. 268–296.

[150] Yanliu Huang, Zhen Yang, and Vicki G Morwitz. “How using a paper versus mobile calendar influences everyday planning and plan fulfillment”. In: *Journal of consumer psychology* 33.1 (2023), pp. 115–122.

[151] Ian Li, Anind Dey, and Jodi Forlizzi. “A stage-based model of personal informatics systems”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2010, pp. 557–566.

[152] Donghan Hu and Sang Won Lee. “Exploring the Effectiveness of Time-lapse Screen Recording for Self-Reflection in Work Context”. In: *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 2024, pp. 1–14.

[153] Eun Kyoung Choe et al. “Understanding quantified-selfers’ practices in collecting and exploring personal data”. In: *Proceedings of the SIGCHI conference on human factors in computing systems*. 2014, pp. 1143–1152.

[154] Dezhi Wu and Marilyn Tremaine. “Knowledge worker adoption of time management tools: Satisfaction and perceived effectiveness”. In: (2004).

[155] Piers Steel. “The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure.” In: *Psychological bulletin* 133.1 (2007), p. 65.

[156] Rowanne Fleck and Geraldine Fitzpatrick. “Reflecting on reflection: framing a design landscape”. In: *Proceedings of the 22nd conference of the computer-human interaction special interest group of australia on computer-human interaction*. 2010, pp. 216–223.

[157] Eric PS Baumer. “Reflective informatics: conceptual dimensions for designing technologies of reflection”. In: *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 2015, pp. 585–594.

[158] Kristina Höök and Jonas Löwgren. “Strong concepts: Intermediate-level knowledge in interaction design research”. In: *ACM Transactions on Computer-Human Interaction (TOCHI)* 19.3 (2012), pp. 1–18.

[159] Stephen J Payne. “Understanding calendar use”. In: *Human-Computer Interaction* 8.2 (1993), pp. 83–100.

[160] Alex C Williams et al. “Managing tasks across the work–life boundary: Opportunities, challenges, and directions”. In: *ACM Transactions on Computer-Human Interaction* 30.3 (2023), pp. 1–31.

[161] Andre N Meyer et al. “’Stick to’Three: Fostering Awareness, Intentions, and Reflections on the Top Daily Tasks”. In: *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 2025, pp. 3647–3662.

[162] Yoana Ahmetoglu et al. “Bridging Analog and Digital: Exploring the Present and Future of Digital Pens in Supporting Planning at Work”. In: *Proceedings of the 4th Annual Symposium on Human-Computer Interaction for Work*. 2025, pp. 1–15.

[163] Meredith Ringel Morris. “Scientists’ Perspectives on the Potential for Generative AI in their Fields”. In: *arXiv preprint arXiv:2304.01420* (2023).

[164] Viktoria Stray and Nils Brede Moe. “Understanding coordination in global software engineering: A mixed-methods study on the use of meetings and Slack”. In: *Journal of Systems and Software* 170 (2020), p. 110717.

[165] Fernanda S Rossi et al. “From glitter to gold: recommendations for effective dashboards from design through sustainment”. In: *Implementation Science* 20.1 (2025), p. 16.

[166] Abraham Mhaidli and Kat Roemmich. “Overworking in HCI: A Reflection on Why We Are Burned Out, Stressed, and Out of Control; and What We Can Do About It”. In: *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. CHI EA ’24. Honolulu, HI, USA: Association for Computing Machinery, 2024. ISBN: 9798400703317. DOI: 10.1145/3613905.3644052. URL: https://doi.org/10.1145/3613905.3644052.

[167] Advait Sarkar. “AI Should Challenge, Not Obey”. In: *Communications of the ACM* 67.10 (2024), pp. 18–21.

[168] Ianina Scheuch et al. “Resilience training programs in organizational contexts: A scoping review”. In: *Frontiers in Psychology* 12 (2021), p. 733036.

[169] Gerben Tolkamp et al. “Creativity as it unfolds: An examination of temporality in the creative process”. In: *Creativity Research Journal* 37.1 (2025), pp. 22–43.