PhD proposal: Predictive modelling of individual human reasoning using belief revision

Clayton Kevin Baker[0000−0002−3157−9989]

University of Cape Town and CAIR
bkrcla003@myuct.ac.za

Abstract. The treatment of old and new beliefs upon learning new information is studied broadly under the term belief change. Many forms of belief change exist e.g. contraction, revision, update, erasure, merging, and fusion. The relationship between belief revision and human reasoning is of primary interest in this work. The theory of belief revision extends classical two-valued logic with an approach to resolve the conflict between a set of beliefs and newly learned information. The goal of this project is to test how humans revise conflicting beliefs. Experiments are proposed in which human subjects are required to resolve conflicting beliefs using two approaches: filtering relevant beliefs from non-relevant ones and discarding beliefs with low confidence. In our analysis, the human responses will be evaluated against the predictions of two perspectives of propositional belief revision: formal and psychological. For the formal aspect, we use the AGM postulates. For the psychological aspect, we use a cognitive model based on Mental Models Theory (MMT). To aid in the predictive analysis for the latter, we will use an open-source program called Cognitive Computation for Behavioural Analysis (CCOBRA). Another goal of this project is to find useful and natural features of human belief revision. We propose to develop an enhanced MMT model that includes these features and link it with CCOBRA.

1 Project description

Reasoning is an activity that is performed daily by human beings. The field of knowledge representation and reasoning studies the abstract task of encoding facts about the world and deriving inferences from it. In cognitive psychology, reasoning is studied by observing patterns of inference directly from human beings. These two fields have interacted increasingly to produce human-like reasoning capabilities in modern technology. Johnson-Laird, a philosopher of language and reasoning, holds the view that humans exploit all the knowledge available to them to arrive at an inference, rather than apply formal logical rules. In response, the task of making formal rules of inference flexible to exceptions and tolerant of new information has been studied under the term non-monotonic reasoning. Oaksford and Chater [30], a proponent of non-monotonic reasoning, view reasoning as a social activity which, when formalised in classical two-valued (true/false) logic, is weakened by a lack of contextual knowledge and the limitation that background knowledge must be consistent. The treatment of old and new beliefs upon learning new information is also studied broadly under the term belief change. Dubois et al. [9] explain this as the process of transitioning from one set of beliefs to another when presented with new information. While many forms of belief change have been studied, e.g. contraction and revision [1], update and erasure [24], merging [27] and fusion [23], the relationship between human reasoning and belief revision is of primary interest in this work. The theory of belief revision extends classical two-valued logic with an approach to resolving conflict between a set of beliefs and newly learned information. For example, a doctor may prescribe a sling, ice, and rest for a patient presenting with acute arm pain after a sports match. During the examination, a doctor may take an x-ray of the patient's arm. If the x-ray reveals a severe fracture, the doctor will be persuaded to realign the arm bones through surgery rather than apply a sling. The overarching objective of this project is to understand more about how humans revise their beliefs when presented with conflicting information. It is proposed that features of human belief revision that are natural and useful be incorporated into contemporary behavioural analysis models.

2 Problem statement

In previous work [3], we surveyed English translations of the 8 Alchourrón, Gärdenfors and Makinson (AGM) [1] postulates of belief revision with human subjects, and half of the postulates were found agreeable with the subjects. A limitation of this work is that only one translation per postulate was used and that a comparison of the data for each postulate was not telling. Since humans are known to construct visual representations, or mental models [20], of their environment, it would be necessary to test each postulate repeatedly using varied material. Another limitation is that although the AGM postulates are a sound framework to examine the relationship between human reasoning and belief revision, the postulates are not considered complete. In recent literature, the AGM theory has been developed through the notions of relevance and independence, in which only the part of a belief set that is affected by the new information should be revised. Syntax splitting, first introduced by Parikh and Chopra [31], considers this key idea concerning revising belief sets. In later work, Aravanis et al. [2] specified epistemic entrenchment [17] and partial meet [1] models for Parikh's relevance-sensitive axiom. Identified as a desirable property of knowledge representation and reasoning, syntax splitting has also been extended to the settings of iterated revision [26], non-monotonic reasoning [25] and contraction [18], amongst others. While the revision theory has been extensively developed for enhancing logical inference, it is less clear whether humans find revision natural and useful.

2.1 Aims

We aim to collect data about how humans handle conflicting beliefs when learning new information. We will compare the responses to formal and psychological accounts of belief revision. The responses will be analysed at the individual and sample levels. For the individual level, we aim to use predictive modelling and analysis. For the sample level, we aim to use statistical modelling and analysis. In our findings, we will discuss whether belief revision is natural and useful in human reasoning. We will also discuss which features of belief revision should be incorporated into decision-making systems.

2.2 Research questions

1. Which features of AGM belief revision do humans use to resolve conflicting beliefs in everyday reasoning?

When incorporating new information into an existing belief set, a revision operation ranks the set of interpretations using total pre-orders [7]. The interpretations (beliefs) that are closest to the old set of beliefs are included in the revised set. This question compares human reasoning to a revision operation. It asks whether humans find consistent beliefs important, and how conflicting information is resolved. The kinds of beliefs that human subjects retain or discard when learning new information should be studied, alongside the context of the situation.

2. Which account of reasoning, AGM belief revision or the Mental Models Theory, predicts human responses to belief revision tasks more accurately?

AGM belief revision is a formal account of reasoning that characterises a revision operation for logical inference, typically in the propositional setting. The AGM approach uses sets of beliefs from which infinitely many consequences, or entailed beliefs, can be generated by applying rules of inference. More specifically, there are eight postulates in the AGM framework that are considered the basis for which a revision system should be judged against: inclusion, closure, success, vacuity, consistency, extensionality, super-expansion and sub-expansion. On the other hand, the theory of mental models or mental models theory (MMT) [19], does not rely on logic alone. Instead, it refers to the semantic cognitive (human) process of representing in memory the available information at hand and selecting a representation which best matches the evidence. The MMT has been applied to propositional reasoning [21] which describes how humans make deductions based on propositions, e.g "the car is parked outside", using *if*, *and*, *or* and *not*. This question asks how the two accounts of propositional reasoning, the formal and the psychological, differ. This question also asks which account is more accurate in predicting human responses to belief revision problems.

2.3 Requirements

The theory of belief revision involves a representation aspect and a reasoning aspect. Generally, beliefs are represented as a set of symbols and reasoning is enabled by manipulating symbols with some underlying logic. The choice of logic depends on how powerful the logic language should be. Classical propositional logic is weak in terms of expressive power since it allows each statement about the world only two outcomes: true or false. However, we will use classical propositional logic to enable comparison with the propositional MMT. Reasoning is typically tested by observing how a subject draws a conclusion from a set of information. For this project, we apply the same principle. A set of reasoning examples is required on which to make inferences, and these examples need to be presented to human subjects in a natural setting. As a result, a core requirement of this work is to collect data from human subjects. Each example should therefore have a diverse set of beliefs that reflect the views of a typical human reasoner. In terms of predicting responses, the material in each example needs to be formatted such that propositions and logical operators are clearly demarcated. To assist with predictive modelling and analysis, we will use an open-source software platform called Cognitive Computation for Behavioural Reasoning Analysis (CCOBRA) [1]. Once tested for propositional logic, the results can be lifted to other forms of logic.

3 Procedures and methods

In the previous section, we have established that a core requirement in answering our research questions involves collecting data from human subjects in an experimental study. We propose surveys to collect data about how human revise their beliefs and deal with conflicting information.

3.1 Experiments

We propose surveys to collect data about how human revise their beliefs. We will recruit undergraduate students at the University of Cape Town to participate in our surveys. The students will be selected across different academic disciplines (humanities, science, engineering, health science, commerce and law). Demographic data such as age, gender, and level of study, will be required. The data will be used to describe the profile of each subject and match the profile with response patterns we identify during the analysis phase. In addition, we will not exclude subjects with prior exposure to logic. However, it will be a requirement that prior experience must be disclosed to account for possible bias in the results. To determine an appropriate a priori sample size, we consider the significance level, power and the treatment groups under observation. The observation that no significant difference exists between two specified groups is called a null hypothesis. Significance (typically 5%) and power (typically 80%)

[1] https://orca.informatik.uni-freiburg.de/ccobra/

[4] refers to the probability of rejecting a null hypothesis when it is true and false respectively. We consider as variables participants' belief relevance and belief confidence. We aim to measure their interaction, if present. Using the default settings on G*Power [14] for a two-way ANOVA interaction test (f = 0.25, $\alpha = 0.05$, df= 10), and setting the power to 0.8, the number of participants needed is 269.

Subjects will respond to a series of conversations between 2 agents. The conversations will refer to places, objects and experiences that humans may encounter in everyday life. The first task (relevance) requires subjects to complete a conversation by selecting and removing statements from a pre-populated belief bank. To select a statement for this task, subjects will assign the statement a value of true. To remove a statement, subjects will assign the statement a value of false. The second task (confidence) requires subjects to analyse a conversation by assigning a confidence estimate (%) to each statement.

3.2 Methods of analysis

The data from both surveys will be combined to produce a full representation of the subjects' beliefs. The variables of interest in the experiment are belief relevance and belief confidence. For each conversation, we will build two corresponding representations. The first depicts the average subject's beliefs w.r.t. relevance. The second corresponds to the average subject's beliefs w.r.t. confidence. A Wilcoxon signed-rank test [8] will be used to measure the difference between relevance and confidence responses. An f-test [6] will be used to measure the difference in sample variance for relevance and confidence responses, while a t-test [29] will be used to measure the difference in sample means. In the next part of the analysis, we want to determine whether MMT can predict individual responses. Conveniently, MMT is a built-in model of CCOBRA [5]. It works by supplying a formatted dataset to CCOBRA and selecting MMT as a parameter. The inferential power of the MMT, via CCOBRA, will be compared to the logical inference of an augmented set of revision postulates. The augmented set will comprise the 8 AGM [1] postulates and desirable postulates for information relevance [32]. For the analysis of individual responses, we will be guided by CCOBRA in terms of response accuracy and subject performance.

3.3 Theoretical contributions

This project builds on previous work that investigates the relationship between belief revision and human reasoning. While the focus of previous work [3] was a litmus test of the AGM postulates with human reasoners, this work considers a broader view of revision. The results of this work will establish the features of belief revision that are natural and useful to human reasoners. Furthermore, this work applies a new methodology to the theoretical domain of knowledge representation and reasoning. This methodology can be reproduced for other forms of logic where human testing is required. A long-term goal of this work is to incorporate useful features of revision in the existing MMT model on CCOBRA.

4 Ethical issues

The ethical issues for this project mainly relate to the management of data, human subjects, and intellectual property. Our strategy for data design, collection, storage, analysis, and distribution is recorded in a data management plan (`https://tinyurl.com/mdp99pe7`). Our strategy for the management of human subjects aligns with the UCT Research Ethics Code for Research Involving Human Subjects (`https://tinyurl.com/2t7c4t8e`), since the experiment sample comprises undergraduate students from the University of Cape Town. As a requirement of this code, we will apply for ethical clearance from the Faculty of Science Human Research Ethics Committee. Ethical clearance will be sought before conducting experiments with human subjects. We aim to compensate each student in our sample commensurate with double the minimum wage rate for metropolitan councils, R27,97/hour [13]. In terms of intellectual property, we commit to the code outlined by the UCT IP Policy (`https://tinyurl.com/ms82kt6r`) which recognises that copyright in publications is automatically assigned by UCT to the author, and in particular, a student owns the copyright in their thesis or dissertation. We note that the potential IP resulting from this work has also not been assigned to any funder. Should any IP materials be developed in the course of this work, the necessary information will be disclosed to the Research Contracts and Innovation department at UCT within 90 days of discovery, by completion of an IP Disclosure Form (`https://tinyurl.com/mpdny557`). For all other professional issues, the UCT Computer Science Graduate Handbook 2022 (`https://tinyurl.com/4dm75yz6`) will be consulted.

5 Related work

Intelligent, automated cognitive reasoning is still a futuristic goal. Johnson-Laird [22] describes the cause as a separation between formal and semantic systems of reasoning. Recently, the platform GPT-3 (Generative Pre-trained Transformer version 3) emerged in the AI community as a means to display current computational intelligence in solving goal-based problems using deep learning. Floridi and Chiriatti [15] found that GPT-3 has a syntactic ability to associate words, but cannot interpret the semantics or context of the request. Studies that test how humans reason are still necessary to reduce the gap between human cognition and automated reasoning. Elio and Pelletier [10] used experiments to test Lifschitz's [28] benchmark problems for non-monotonic reasoning, asking what follows from a set of premises embedded in a cover story. Using a similar approach, Ford [16] established that humans find non-monotonic reasoning difficult when the conclusion of an argument differs in specificity [34] and logical strength. Elio and Pelletier's experimental testing of default reasoning [12] found that natural language problems were preferred by human subjects over artificial ones, and that information about class size and inter-object similarity yielded stronger inferences. In recent work, the focus has shifted towards combining theoretical and experimental approaches. Elio and Pelletier [11] used experiments to study how logical syntax influenced revision and update. The logical representations used were conditionals, its converse, biconditionals, and two logical rules of inference: modus ponens and modus tollens. Conditionals, or conditional statements, are present in everyday reasoning. For example, "if it is sunny, then it is not rainy". A conditional has the general form *if* antecedent *then* consequent. Modus ponens and modus tollens are logical rules of inference. Modus ponens states that if a conditional statement is accepted, and the antecedent holds, then the consequent may be inferred. Cramer et al. [5] also used experiments to compare human performance on conditional statements with weak completion semantics, a three-valued theory of non-monotonic reasoning. Predictive analysis, a technique in which cognitive models of reasoning are embedded with machine learning to identify individual responses in a dataset, has been applied by Riesterer et al. [33] in the domain of syllogistic reasoning. However, the predictions of the syllogistic model were found inferior to simpler frequency analysis. Efforts to pursue the simulation of cognitive reasoning are thus important, particularly those that study what conclusions humans make together with how humans arrive at them

6 Anticipated outcomes

In the following, we discuss the novel artefacts, expected impact and key success factors that shape the outcome of this project.

6.1 Novel artefacts

A goal of this work is to establish features of belief revision that are useful and natural for humans. To do this, we first collect data about how human subjects revise their beliefs in everyday settings. The inferences from human subjects will be evaluated against the logical inferences of postulates and properties of belief revision, as well as the MMT. If the predictive analysis of belief revision performs shows better results than the MMT, we propose to develop an MMT cognitive model enriched with significant features of human belief revision. The model will be deployed on CCOBRA, and documentation will be provided for its use. In addition, there is potential to lift these results to higher-order logics. However, if belief revision performs no better than the MMT, we will discuss the current limitations of both models and trial new experiments to determine specific domains to which each model is better suited.

6.2 Expected impact

Previous work has tested the AGM postulates of belief revision with human reasoners in a natural language setting. The results indicated that nearly half of the postulates were applied in the reasoning examples. This shows that belief revision is used to some degree in human reasoning. However, the AGM postulates alone are not complete in their ability to deal with conflicting information. This work proposes to investigate belief revision with human reasoners in a broader context. The results from this project will be stronger since we triangulate a statistical analysis of the sample data with predictive analysis of individual responses. Part of this project includes developing a set of reasoning examples to test belief revision. As a consequence, these examples can serve as a benchmark for testing other non-AGM models of revision with human reasoners. Finally, the results of the project will contribute towards the greater goal of modelling human cognition.

6.3 Key success factors

Success is defined as achieving a specific project task, deliverable, or milestone within a specific timeline that positively impacts the project. We divide the project timeline into yearly work units. For the first year of this project, the pertinent success factors are:

- a rigorous analysis of the AGM postulates, its limitations, and improvements
- a rigorous analysis of the MMT, its limitations, and improvements
- a prototype of the experiment design with natural language examples to test human belief revision

For the second year, the success factors are:

- a finalised experiment design with natural language examples to test human belief revision
- a reproducible statistical analysis of the sample responses
- a reproducible predictive analysis of the individual responses
- the identification of features of human belief revision
- a prototype of an MMT model enhanced with features of human belief revision

For the final year of this project, the success factors are:

- a finalised prototype of an MMT model enhanced with features of human belief revision
- testing and deployment of the enhanced MMT model on CCOBRA
- documentation of the enhanced MMT model

7 Project plan

In the following, we discuss the project timeline, risks, resources required, deliverables, and milestones.

7.1 Timeline

This project spans 3 years, from June 2022 to June 2025. The project is currently near the end of the first year. A Gantt chart is provided in Appendix A for each year, 1 - 3, of the project.

7.2 Risks

A risk matrix that compares each project risk with its impact, probability, and management strategies is provided in Appendix B. 14 project risks have been identified. The risks range from low to high both in impact and probability. However, all risks shown can be mitigated.

7.3 Resources required

This project involves literature that spans two distinct disciplines, knowledge representation and reasoning, and cognitive psychology. The primary supervisor, Prof. Thomas Meyer, will provide expertise and guidance on the former aspect, while an established project collaborator, Prof. Marco Ragni from TU Chemnitz, will provide the expertise and guidance on the latter. In terms of physical resources, access to the internet and a standard laptop is required. A Python compiler is needed to run CCOBRA, a software platform used for part of our data analysis.

7.4 Deliverables

The following deliverables are units of physical or digital work. Each deliverable is a result of completing a set of project tasks:

- a project proposal document
- a project proposal presentation
- surveys that test human belief revision
- a software prototype that combines MMT with useful features of human belief revision and executes on CCOBRA
- a Ph.D. thesis

7.5 Milestones

The completion of each project year is viewed as a project milestone. In year 1 (Figure 1), the focus is on developing a sound literature review and proposal. In year 2, (Figure 2), we will develop new software and theory artefacts through experiments. In the last year, (Figure 3), the focus is to document new artefacts and produce a Ph.D. thesis which concludes this project.

References

1. Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. Journal of Symbolic Logic **50**, 510–530 (1985). `https://doi.org/10.2307/2274239`
2. Aravanis, T., Peppas, P., Williams, M.A.: Full characterization of parikh's relevance-sensitive axiom for belief revision. Journal of Artificial Intelligence Research **66**, 765–792 (2019)

3. Baker, C., Meyer, T.: Belief change in human reasoning: An empirical investigation on mturk. In: Proceedings of the Second Southern African Conference for Artificial Intelligence Research (SACAIR 2021). pp. 520–536 (2021)
4. Brysbaert, M.: How many participants do we have to include in properly powered experiments? a tutorial of power analysis with reference tables. Journal of cognition (2019)
5. Cramer, M., Hölldobler, S., Ragni, M.: Modeling human reasoning about conditionals. In: 19 th International Workshop on Non-Monotonic Reasoning. p. 223 (2021)
6. Delacre, M., Leys, C., Mora, Y.L., Lakens, D.: Taking parametric assumptions seriously: Arguments for the use of welch's f-test instead of the classical f-test in one-way anova. International Review of Social Psychology **32**(1) (2019)
7. Delgrande, J.P., Peppas, P., Woltran, S.: General belief revision. Journal of the ACM (JACM) **65**(5), 1–34 (2018)
8. van Doorn, J., Ly, A., Marsman, M., Wagenmakers, E.J.: Bayesian rank-based hypothesis testing for the rank sum test, the signed rank test, and spearman's ρ. Journal of Applied Statistics **47**(16), 2984–3006 (2020)
9. Dubois, D., Everaere, P., Konieczny, S., Papini, O.: Main issues in belief revision, belief merging and information fusion. A Guided Tour of Artificial Intelligence Research: Volume I: Knowledge Representation, Reasoning and Learning pp. 441–485 (2020)
10. Elio, R., Pelletier, F.J.: Human benchmarks on ai's benchmark problems. In: Proceedings of the fifteenth annual conference of the cognitive science society. pp. 406–411 (1993)
11. Elio, R., Pelletier, F.J.: The effect of syntactic form on simple belief revisions and updates. In: Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society. pp. 260–265. Routledge (2019)
12. Elio, R., Pelletier, F.J.: On reasoning with default rules and exceptions. In: Proceedings of the eighteenth annual conference of the cognitive science society. pp. 131–136. Routledge (2019)
13. of Employment, D., Labour, G.o.S.A.: Minister thulas nxesi hikes minimum wage of farmworkers and domestic workers. Online (February 2023), `https://www.gov.za/speeches/minister-thulas-nxesi-hikes-minimum-wage-farmworkers-and-domestic-workers-22-feb-2023-0000`
14. Faul, F., Erdfelder, E., Lang, A.G., Buchner, A.: G* power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods **39**(2), 175–191 (2007)
15. Floridi, L., Chiriatti, M.: Gpt-3: Its nature, scope, limits, and consequences. Minds and Machines **30**, 681–694 (2020)
16. Ford, M.: Human nonmonotonic reasoning: The importance of seeing the logical strength of arguments. Synthese **146**(1), 71–92 (2005)
17. Gärdenfors, P., Makinson, D.: Revisions of knowledge systems using epistemic entrenchment. In: Proceedings of the 2nd conference on Theoretical aspects of reasoning about knowledge. pp. 83–95 (1988)
18. Haldimann, J.P., Kern-Isberner, G., Beierle, C.: Syntax splitting for iterated contractions. In: Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning. vol. 17, pp. 465–475 (2020)
19. Johnson-Laird, P.N.: Mental models. (1989)
20. Johnson-Laird, P.N.: Mental models and human reasoning. Proceedings of the National Academy of Sciences **107**(43), 18243–18250 (2010)

21. Johnson-Laird, P.N., Byrne, R.M., Schaeken, W.: Propositional reasoning by model. Psychological review **99**(3), 418 (1992)
22. Johnson-Laird, P.: Possibilities and human reasoning. Possibility Studies & Society p. 27538699231152731 (2023)
23. Jøsang, A., Costa, P.C., Blasch, E.: Determining model correctness for situations of belief fusion. In: Proceedings of the 16th International Conference on Information Fusion. pp. 1886–1893. IEEE (2013)
24. Katsuno, H., Mendelzon, A.O.: On the difference between updating a knowledge base and revising it. Belief revision p. 183 (1991)
25. Kern-Isberner, G., Beierle, C., Brewka, G.: Syntax splitting= relevance+ independence: New postulates for nonmonotonic reasoning from conditional belief bases. In: Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning. vol. 17, pp. 560–571 (2020)
26. Kern-Isberner, G., Brewka, G.: Strong syntax splitting for iterated belief revision. In: IJCAI. pp. 1131–1137 (2017)
27. Konieczny, S., Pino Pérez, R.: Logic based merging. Journal of Philosophical Logic **40**, 239–270 (2011)
28. Lifschitz, V.: Benchmark problems for formal nonmonotonic reasoning. In: Reinfrank, M., de Kleer, J., Ginsberg, M.L., Sandewall, E. (eds.) Non-Monotonic Reasoning. pp. 202–219. Springer Berlin Heidelberg, Berlin, Heidelberg (1989)
29. Liu, Q., Wang, L.: t-test and anova for data with ceiling and/or floor effects. Behavior Research Methods **53**(1), 264–277 (2021)
30. Oaksford, M., Chater, N.: New paradigms in the psychology of reasoning. Annual review of psychology **71**, 305–330 (2020)
31. Parikh, R.: Beliefs, belief revision, and splitting languages. Logic, language and computation **2**(96), 266–268 (1999)
32. Peppas, P., Williams, M.A.: Kinetic consistency and relevance in belief revision. In: Logics in Artificial Intelligence: 15th European Conference, JELIA 2016, Larnaca, Cyprus, November 9-11, 2016, Proceedings 15. pp. 401–414. Springer (2016)
33. Riesterer, N., Brand, D., Ragni, M.: Predictive modeling of individual human cognition: Upper bounds and a new perspective on performance. Topics in cognitive science **12**(3), 960–974 (2020)
34. Touretzky, D.S.: Implicit ordering of defaults in inheritance systems. In: AAAI. pp. 322–325 (1984)

A Gantt chart

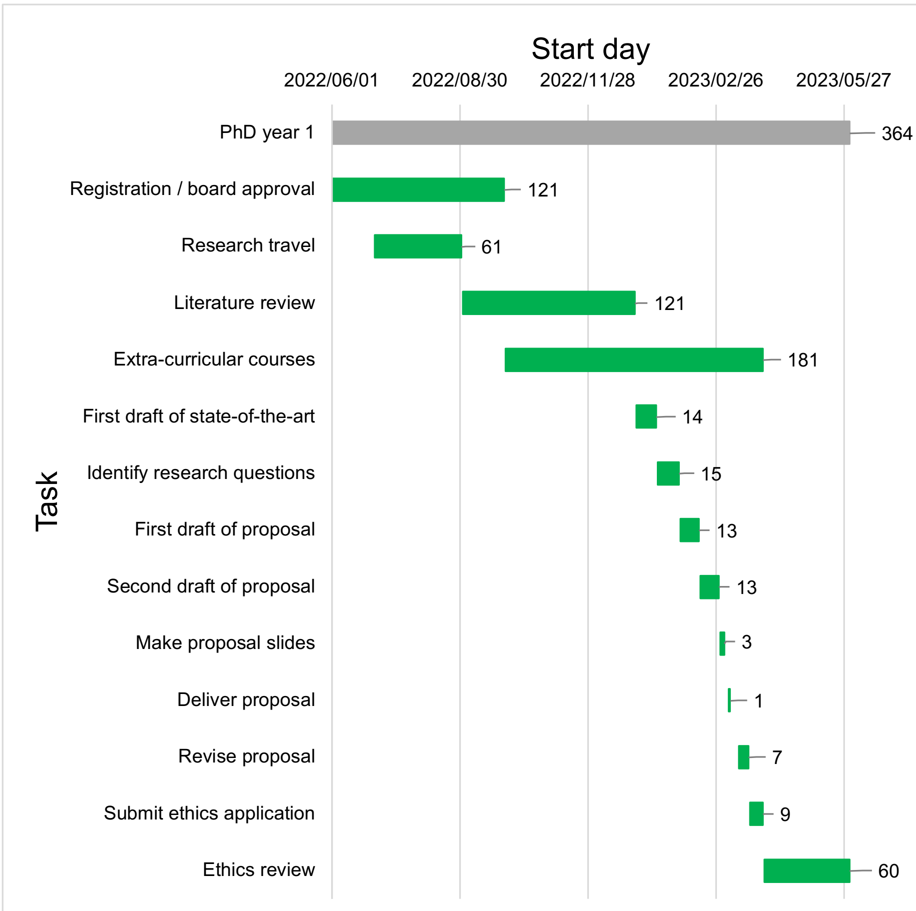

Fig. 1. Gantt chart showing task name versus start date for PhD year 1

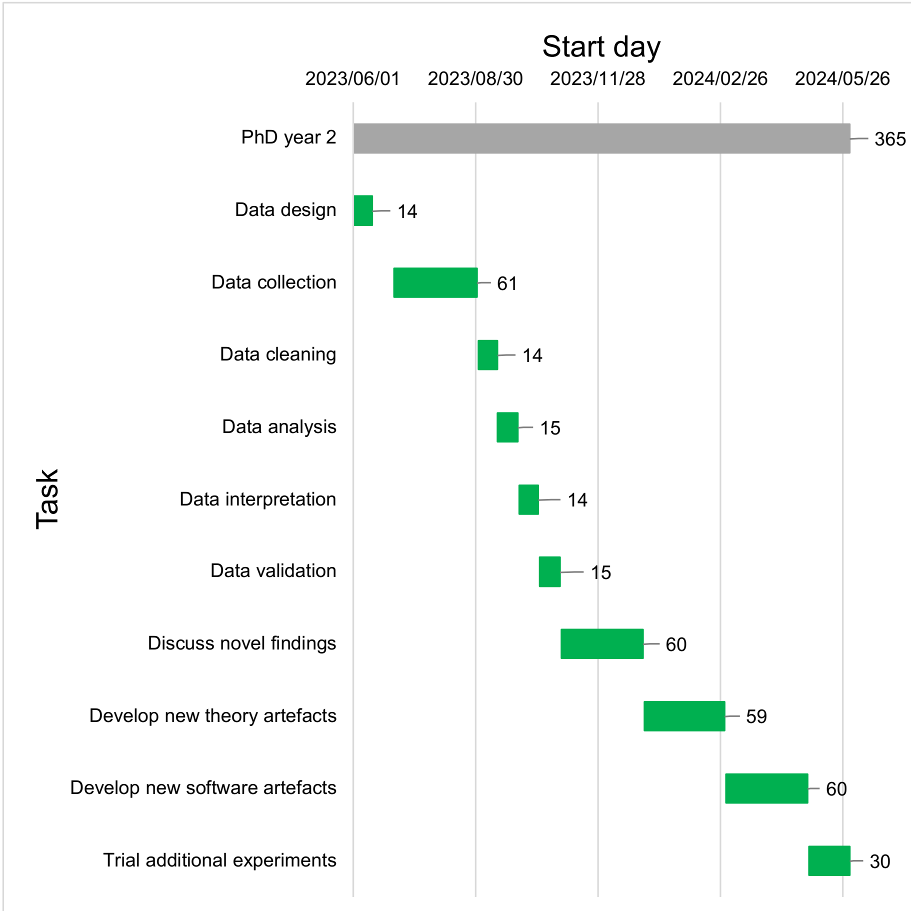

Fig. 2. Gantt chart showing task name versus start date for PhD year 2

B Risk matrix

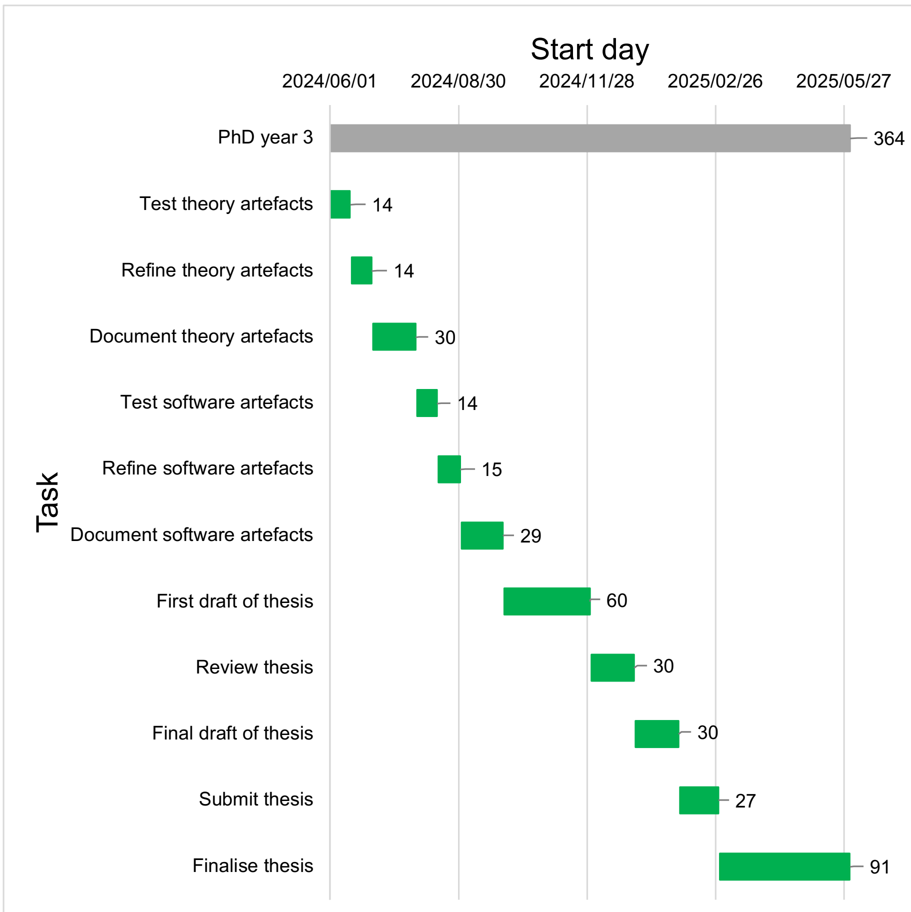

Fig. 3. Gantt chart showing task name versus start date for PhD year 3

No.	Risk description	Impact	Probability	Management/Mitigation strategy
1	Scope creep	high	high	The project scope must be clearly defined in the project proposal document. After the proposal has been presented, the scope should be adjusted to reflect the feedback from the proposal committee. Minor changes can be implemented by the researcher with oversight from the supervisor.
2	The thesis exceeds the maximum word limit	low	moderate	The thesis should be written in stages. After each completed stage, the writing should be reviewed by the researcher and the supervisor. Supplementary information can be included in an appendix. The thesis body should be concise, and additional work outside the project scope can be considered for publication.
3	Research becomes outdated/stale	high	moderate	The researcher should stay abreast of the literature by engaging in active literature review. The researcher should meet regularly with the supervisor to discuss issues. The researcher should discuss his work with peers and participate in appropriate seminars, workshops and conferences.
4	Data are not reproducible	high	low	The experiment and data design must be clearly defined. The ethics application documentand data management plan provides guidance for creating, storing and distributing data.
5	Selection/sampling bias	high	low	Participants should be selected in a fair and transparent manner. Participants should provide informed consent and may exercise their right to remove consent. The sample size should be determined from the variables under investigation and the type of analysis that will be performed. Experiment material should be presented in
6	Researcher confirmation/reporting bias	high	low	The raw data should be published alongside the interpreted data. The results should be triangulated with different forms of analysis. The results and analysis should be published for peer review.
7	Researcher falls ill/becomes incapacitated	high	low	For mild illness, the researcher should rest and resume work once possible. For severe illness, the researcher should apply for sick leave and resume work once possible. If the registration period has exceeded 3 years, the researcher should apply for an extension. In case of incapacitation, the PhD project will be terminated.
8	Results are not significant/novel	moderate	moderate	Review methodology and experiment setup. Relate findings to existing theory/software artefacts. Develop new artefacts. Trial additional experiments.
9	Irregular access to campus due to protests and	low	low	Where meetings are not possible on campus, make alternative arrangements (online, or off-site venue).
10	Irregular internet access due to load-shedding	high	high	The researcher will invest in a UPS for remote online work. The researcher should, in additional to a physical copy, use a backup service to store an online copy of the research files.
11	Dispute between supervisor and researcher	high	low	The researcher should attempt to resolve the dispute amicably. Otherwise, the researcher should follow the channels in the MoU/PPA to resolve the dispute with the supervisor.
12	Shortage of research funds	moderate	low	The researcher should prepare a budget for research spending in the proposal document. The researcher should discuss the budget and appropriate funding channels with the supervisor.
13	Data leak/participant privacy	low	low	The risk to participant privacy is low since the material in the experiments refer to impersonal situations, objects and places. However, to respect participant privacy, datasets will be stripped of personal identifiers before being published.
14	Stress, poor work/life balance	high	high	The researcher should take regular breaks from periods of work, follow a balanced diet and do physical activity. The researcher should stay in regular communication with the supervisor.

Fig. 4. Matrix of risk, impact, probability and management strategy