

Aspectually Enhanced Modelling of Sociotechnical Systems

Tom Wallis

Submitted in fulfilment of the requirements for the Degree of Doctor of Philosophy

School of Computing Science

College of Science and Engineering

University of Glasgow

© Tom Wallis

Table of Contents

1	Introduction	27
1.1	A primer on aspect orientation	28
1.2	Prior Work	28
1.3	Terms & definitions	29
2	Relevant Literature	31
2.0.1	Motivations & Philosophy Underlying Aspect Orientation	32
2.0.2	Criticisms of Aspect Orientation	36
2.0.3	Alternatives to Aspect Orientation	38
2.0.4	Subject-oriented Programming	40
2.1	Dynamic Aspect Weaving	41
2.1.1	Approaches to Dynamic Weaving	41
2.1.2	PROSE	42
2.1.3	Nu	43
2.1.4	Dynamic Weaving in Embedded Systems	44
2.2	Aspect Orientation in Simulation & Modelling	45

2.2.1	Suitability of Aspect Orientation in Experimental Codebases	45
2.2.2	Aspect Orientation in Discrete Event Simulation	47
2.2.3	Aspect Orientation & Business Process Modelling	50
2.3	Process Variance in Simulation & Data Generation	53
2.3.1	Discussion of Variation & Motivations for Variations in Process Models	54
2.3.2	Representing Variations in Process Models and their Outputs	55
2.4	Discussion	58
2.4.1	Research Questions	61
3	Prior Work	63
3.1	Motivations in originally implementing PyDySoFu	63
3.2	Original PyDySoFu Implementation	65
3.2.1	Weaving mechanism	65
3.2.2	Applying Process Mutations	67
3.2.3	Limitations	67
3.3	Additional Simulation Machinery	70
3.3.1	Fuzzi-Moss	70
3.3.2	Theatre_Ag	72
3.4	Example Studies using PyDySoFu for Behavioural Simulation	73
3.5	Discussion	74
4	Rewriting PyDySoFu	77
4.1	Requirements for Change	77

<i>TABLE OF CONTENTS</i>	<i>5</i>
4.2 Python3-compatible Weaving Techniques	79
4.2.1 Abandoned techniques	79
4.2.2 A viable technique: import hooks	80
4.3 Import Hooks	82
4.3.1 Implementing import hooks	82
4.3.2 Strengths and weaknesses of import hooks	83
4.3.3 Weaving process	85
4.4 Optimisations	89
4.4.1 Deep Hook Weaving	90
4.4.2 Non-Dynamic Weaving	90
4.4.3 Priority Ordering of Aspect Application	91
4.5 Discussion	92
5 RPGLite: A Mobile Game for Collecting Data	93
5.1 An Overview of RPGLite	95
5.1.1 RPGLite’s Design Implications	96
5.2 Implementation of RPGLite	99
5.2.1 Consent to participation	99
5.2.2 Mobile app	100
5.2.3 API & Server-side Logic	102
5.3 Empirical Play and Data Collected	103
6 Simulation Optimisation with Aspect Orientation	105

6.1	Aims	106
6.1.1	PyDySoFu Suitability	108
6.1.2	Proposed Experiment	109
6.2	Naive Model	111
6.3	Experimental Design	114
6.3.1	Models of Learning	116
6.3.2	Modelling Learning in RPGLite	118
6.3.3	Modelling Player Confidence in RPGLite	121
6.4	Aspects Applied	122
6.4.1	Aspects for model improvement	123
6.4.2	Aspects for Instrumentation	126
6.4.3	Aspects Implementing Models of Learning	129
6.5	Experimental Design	132
6.5.1	Optimising the Model to Simulate a Real-World Player	133
6.5.2	Model parameters	133
6.5.3	Representing the result of an experimental run	137
6.5.4	Quantifying Similarity of Character Pair Selection	137
6.5.5	Controlling State Space Exploration	140
6.5.6	Producing Datasets	141
6.5.7	Answering the first Research Question	142
6.6	Experimental Results	142

6.6.1	Identifying Statistically Significant Results in Datasets	143
6.6.2	Selecting Character Pairs from a Known Distribution	144
6.6.3	Selecting Character Pairs using Confidence-Based Learning	145
6.6.4	Learning with Hyperbolic Decay	145
6.7	Discussion	145
7	A chapter title here for the experiment moving aspects to new systems, or systems with some changes	147
7.1	(Reword) Experimental motivations / motivation of research question	147
7.1.1	Coupling of Model and Behaviour	148
7.1.2	(Reword) Gaps/opportunities left by previous experiment	148
7.1.3	Research Question	148
7.2	Experimental Design	148
7.2.1	Changes to RPGLite	148
7.2.2	Applying Behavioural Variations to New System	149
7.3	Applying Aspects from a Control System to a New System	149
7.3.1	Implementation	149
7.3.2	Results	149
7.4	Discussion	149
8	Future Work	151
8.1	Aspect-Oriented Metaprogramming in real-world Software Engineering	152
8.2	Augmentation of pre-existing models	152

8.3	Aspect Orientation's utility for Research Software Engineers	153
8.4	Hypothesising Possible System Dynamics via Aspects	154
8.5	Standards for Aspect Orientation in Research Codebases	157
8.6	Standard Aspect-Oriented Model Features	158
8.7	Testing Frameworks to detect Model Incorrectness	159
8.8	Optimisation of Multiple Aspects	162
8.9	Future Work pertaining to RPGLite	162
8.9.1	Causes of Game Abandonment	163
8.9.2	Patterns of Play within Cliques of Players	163
8.9.3	Aspect Oriented Discovery of Dataset Properties	164
8.9.4	Larger-scale Data Collection	165
8.10	Discussion	166
9	Closing Discussion	169

Todo list

[C1] todo: Explain aspect orientation briefly.	28
[C1] todo: find a citation for spring AOP	28
[C1] todo: Write a section for the introduction describing the work done on pdsf before this, to delineate where we're starting from and avoid any claims of plagiarism. This can be short, the first sec of the lit review is a proper discussion, but the tool should be mentioned here. See ?? for what already exists.	29
[C1] todo: Complete the glossary in section 1.3.	29
[C1] todo: Decide whether terms like BPMN, simulation & modelling, etc also belong in the glossarysection 1.3.	29
[C1] todo: Dejice's thesis has a similar glossary; how did he structure it? Wasn't an appendix. Was it a section of the intro or something else?	29
[C2] review: add citations where code quality improvements are claimed in AOP research	31
[C2] review: There's a paragraph missing here explaining which chapters follow, why, and in what order. I'm omitting it at the moment because I suspect I'm about to reorder them. . .	31
[C2] review: Reword if I move criticisms to another section.	31
[C2] review: I believe this para was mangled by emacs somehow, and is the end of a drafted final paragraph of this introduction. Keeping it because it wasn't committed so can't restore the original para, and ideally I'll recreate it when redrafting this chapter.	31

<i>[C2] todo: Cite some OO languages, papers noting their industrial use / the “success” of object orientation.</i>	32
<i>[C2] review: Feels like the full literature on AOP is missing here. What is the available evidence that it works, if any? What has it been used for? Some sort of survey of AOP tech might also be useful for talking about alternatives?</i>	35
<i>[C2] review response/Q: I wonder whether I need to review more literature (possible — I have a few mentioned in comments here and in annotated.bib that I should look at if I have time, though I believe I’ve covered the essentials) or whether the tech that’s of interest here is reviewed later on. I suspect it’s included later, because there’s a note toward Dynamic Aspect Weaving which suggests the tech needs to come before the criticisms review (i.e. there’s some tooling reviewed, but not where it should be). Mostly folk seem to have stuck with AspectJ, and domain-specific / general tooling is reviewed later. I’m reticent to move the specific tech reviewed here because I currently have a nice separation between the philosophy of the paradigm and its implementations; maybe I should make those concepts more clearly separated in section headings, or maybe I should move the tech here and make Criticisms its own section, rather than a subsec of this introduction.</i>	35
<i>[C2] review: Why? Presumably because they use the wrong type of modularity (Lexical?) — if so, make this clear here!</i>	37
<i>[C2] review: This review is arguably unfinished — there are more papers commented below this note which are probably worth reading / ensuring I’m citing. Tim has noted that the review on early / foundational literature of AOP seems incomplete, and I think the commented papers to review below could be good candidates to address this.</i>	37
<i>[C2] review: this needs more explanation. What were Kiczales et al. trying to do, what was the context, what did they demonstrate?</i>	38

[C2] review response/Q: I covered this a fair bit in the earlier parts of this section. Maybe this was unclear? This lit is already covered, so fixing earlier notes might be better than adding notes here. I've since restructured & reframed, hopefully enough to make this clearer, but I'm curious...	38
[C2] todo: Add notes on Bigraphs here	39
[C2] todo: Add notes on Holons here, maybe	39
[C2] review: Wouldn't the paragraph below fit better in the section on program mutation?	39
[C2] review response/Q: I'm not sure exactly what section this is, but have reworded the below to fit better within this section (BCA is a relative of AOP) and added a cref to where I <i>think</i> it might be added. Unsure though. To discuss and fix the cref. I haven't added notes on BCA anywhere else. Or, was it intended to refer to the metaprogramming notes in the next para? If so, I've since collected these into a subsubsec.	39
[C2] todo: A diagram or Python example might help here, as per Tim's suggestion below.	40
[C2] review: Would some illustrative examples of these different approaches help?	40
[C2] review response/Q: Yes, I'm sure they'd help! Diagrams to add here.	40
[C2] review: Incomplete! Add subject-oriented programming here.	40
[C2] todo: Note: there's an argument that SOP has seen more success in practice than AOP: Golang's interfaces are subjective views on a struct, and I believe Rust's Traits are too — confirm this! — by contrast, there's some limited use of AOP in the wild, but related design patterns have arguably supplanted them for modularising cross-cutting concerns (I'm thinking of Decorators). I'm not aware of any literature which backs this up.	40
[C2] review: I think this comes before the critique section perhaps? Need a general introduction to AOP techniques and techs, including this?	41
[C2] review: T: Have I actually explained what a dynamic weaver is? Think not.	41

[C2] review: T: maybe this is a suitable introduction to dynamic weaving, if I add an explanatory paragraph at the beginning?	41
[C2] review: Need citations to these technologies	41
[C2] review: AspectJ belongs here.	42
[C2] review: I don't define adaptive programming. Maybe I should have it as an alternative? End of this section, after subject-oriented programming?	43
[C2] review: Where's this quote from? Citation needed. Think it could be Przybylek [Prz10] or Steimann [Ste06].	44
[C2] review: T: If there's time when redrafting, note which scenarios they distinguish between specifically.	46
[C2] review: it seems like this might be an important exposition, so perhaps more details?	46
[C2] review response/Q: This is covered in the criticisms of AOP. Maybe I should just make a reference here, or maybe there's room for a paragraph to hammer the point home. Ask Tim which he thinks is more appropriate.	46
[C2] review: Evaluating?	47
[C2] review: new sentence here - Chibani, Belattar, and Bourouis [CBB19]'s later work presented an implementation of their proposal and provide a quantitative evaluation.	47
[C2] review: grab a citation for Simkit	48
[C2] review: add citations where code quality improvements are claimed in AOP research	49
[C2] review: past tense again - the paper has been written/published - foresaw, observed, described... 50	50
[C2] review: what do you mean by static nature? not clear	51
[C2] review: Can you give an example	51
[C2] review: this seems relevant argument so avoid footnotes?	51

TABLE OF CONTENTS	13
[C2] review: Don't need to state pub type, just name the author	51
[C2] review: can you say something explicit about the difference and benefits of this?	52
[C2] review: Can you say something more precise than spirit here, e.g. objectives, intentions at least?	52
[C2] review: demonstrates	52
[C2] review: conventions, syntax - again be more precise	52
[C2] review: add something early on giving an example of how it might be a problem?	53
[C2] todo: Brief intro of the process variance section here	53
[C2] review: Agree, builds well from preceding sections though	53
[C2] review: Not sure this is true - e.g. MAS?	53
[C2] review: remove plausibly - it can or it can't	54
[C2] review: Has this motivation already been explained? If not re-order with below first.	55
[C2] todo: Find citation for empirical datasets being the only ones available to researchers, maybe for disaster prevention or similar?	55
[C2] todo: find a citation for simulating different systems for finding an optimal one. Arguably Genetic programming & hill climbing do similar things?	55
[C2] todo: Find some nice way to round this subsection out, or refactor out subsections	55
[C2] todo: Important to acknowledge somewhere in this subsection that synthetic data generation is a well-researched field, but that generating logs from simulations with variance is the specific area relevant to this thesis	55
[C2] review: refactor by re-ordering - explain the motivation first and then intro the approach of charfi	55

[C2] review: with what particular characteristics? e.g. that plausibly mimic real world logs of real security violations?	56
[C2] review: is this a however, i.e. a limitation?	56
[C2] review: further	56
[C2] review: techniques	56
[C2] review: say limited instead of relatively rudimentary?	56
[C2] review: why only presumably?	56
[C2] review: new paragraph?	56
[C2] review: of?	56
[C2] review: be explicit - on what. Also start with the overall thrust of the work, not the specific contribution	57
[C2] review: as per first comment on this paragraph - what technique?	57
[C2] todo: Shugurov & Mitsyuk's work should follow this. REVIEW IT.	57
[C2] review: unclear here	57
[C2] review: have been developed	58
[C2] review: This sentence was salvaged from the simulation & modelling section. It's OK but was waffly there. Maybe it can fit in this discussion somewhere?	58

[C2] review: Worth observing that Ionescu et al. [Ion+09] actually produce a case study of using AOP and their code quality evaluation uses SLOC (unreliable), but they do find that they can augment models successfully (in their context, successfully is without significant loss of performance). Backs up claim in Gulyás and Kozsik [GK99] that AOP is useful for simulation & modelling and claim in Steimann [Ste06] that AOP “doesn’t know what it aims to be good for” (paraphrased, it’s quoted in the criticisms section and used elsewhere too). There’s loads of AOP lit but we need to explore using it in new ways, and actually building models & running simulations using it can be motivated by existing literature. I don’t believe I make reference to Ionescu et al. [Ion+09] in this section at present.	58
[C2] todo: more citations	59
[C2] review: This is good, but don’t be speculative like this. Draw a conclusion	60
[C3] review: Consider moving some code snippets like those of Fuzzi-Moss, which aren’t necessarily that important in the grand scope of the thesis, to an appendix.	72
[C3] review: Maybe I should include something on Aran’s masters dissertation too, rather than just citing it?	73
[C3] review: Add a citation for TDD vs agile! Citations are missing from the copy of the CAiSE paper I can find, but I’d like to use the same one.	73
[C3] review: Does this want to be a list? Consider reworking to a paragraph.	74
[C4] todo: Check the most up-to-date pdsf implementation — is it actually on that backup drive?	77
[C4] review: Rewrite chapter outline after the structure of the chapter is known to be OK.	77
[C4] todo: Consider adding references to the sections through this PDSF chapter, depending on how beefy it becomes...	77
[C4] review: Too informal	77
[C4] review: sloppily put	77
[C4] review: Could be less sloppy	78

[C4] review: Could be less sloppy	78
[C4] todo: Consider adding local namespace weaving to pdsf3: should be easy to implement as a cheeky little monkey-patch...	83
[C4] todo: is this true??? verify, I think not, if imported normally. tested but didn't realise that pdsf won't fuzz functions with single-character IDs, so inconclusive.	84
[C4] review: A bit clunky — take the last sentence out in favour of notes in discussion or rewrite.	86
[C4] review: Decide whether the “-style” prefix I sometimes use should stay or go when copy- editing. Or: when do I use it? Be consistent.	87
[C4] review: Add metrics here on the performance implications; they're pretty significant, I actually added them because the regex matches slowed my own code down to a crawl. Our experiments make use of non-dynamic weaving for this reason.	91
[C4] review: contributions to what?	92
[C4] review: Is this list complete?	92
[C5] review: About what, and why? Consider changing title. add a sentence in first paragraph describing why this fits in with the rest of the dissertation, and why it is here.	93
[C5] review: Suggest new sentence	93
[C5] review: PRISM undefined	93
[C5] review: “game states” not well defined here, though I do define it later on	93
[C5] todo: Fix formatting and correct wording of William's hypothesis. Review note from Jess: this is <i>very</i> important (and, yes! I don't believe i have my RQs well-defined anywhere in the thesis yet, which is concerning.)	93
[C5] todo: Insert my own research question here	93
[C5] review: Lift screenshots in this chapter from game-on paper source where possible	93
[C5] review: “cost” undefined	93

<i>[C5] review: From Jess: this seems important, but needs clarification and reworking. She suggests moving it to the top.</i>	93
<i>[C5] todo: TODO INSERT RESEARCH QUESTION HERE</i>	93
<i>[C5] review: From Jess: The last half of this para in particular is hard to understand.</i>	93
<i>[C5] review: J: rework passive voice</i>	94
<i>[C5] review: J: this para's easier to understand than previous ones in this chapter. When reworking, consider how the others could be rewritten in a style closer to this? She also notes it might belong further up in the chap.</i>	94
<i>[C5] review: J: avoid passive voice</i>	94
<i>[C5] review: J: She suggests moving this earlier, too.</i>	94
<i>[C5] review: Needs copy-editing</i>	95
<i>[C5] review: J: compress the description list. Maybe make it bulleted? I wonder whether it can just be a para or something.</i>	95
<i>[C5] review: J: whose? the barbarian's? Clarify.</i>	95
<i>[C5] review: Jess noted this wasn't formally described. I describe it later, but clearly it's confusing on a first read; consider either reordering or including a definition earlier, the term shouldn't be used here if it's not defined properly.</i>	96
<i>[C5] review: J: so is that the state, or just a description of a subset? (It's a complete and uniquely identifying state, but I don't make this clear enough)</i>	97
<i>[C5] review: J: define earlier? Also: why is entropy important here?</i>	97
<i>[C5] review: J: does the order matter?</i>	97

[C5] review: J: Summarising her notes but 1. Footnote should be integrated into body and 2. Entropy's a confusing concept, wasn't what she was expecting here for defining a state space. Either make the concept extremely clear or, better, avoid the concept of entropy so as not to trip people up. Clearer writing > clever writing.	97
[C5] review: J: suggest a diagram here	97
[C5] review: J: this needs to be defined too	97
[C5] todo: Metagame is never defined!	97
[C5] review: J: first section isn't clear, she notes she didn't understand. In second section she circles "known", asking, "known by whom?"	98
[C5] review: Footnote suggests correlation is defined in chap 6; include a more specific cref. . .	98
[C5] review: J: compress	99
[C5] review: Should a summary of the paper be a section here, or maybe an appendix? Keen not to plagiarise, but it's also maybe worth including. William and I both spent a lot of time getting that paper out. I presented it at GameOn2020. He's included some in his thesis, so I guess it's a precedent for me to do the same. And why not?	99
[C5] review: Should info sheet / consent form be cited, instead of linked as in this footnote...? .	99
[C5] review: J: rest of this para should be made shorter.	100
[C5] review: There are some good screenshots in the paper for RPGLite; lessons learned. Consider adding those here instead; the screenshots I've included at present come from my iCloud photos library...	100
[C5] review: J: how could the data not be empirical?	102
[C5] review: J: suggests this is separate from games. Is that correct? (Yes; make this clearer.) . .	103
[C5] review: J: make more concise. (I made slight edits when adding this revnote, maybe the para can be condensed. I haven't done much right now.)	103

<i>[C5] review: citation undefined</i>	103
<i>[C5] review: ELO undefined.</i>	104
<i>[C5] review: J: “score” is vague. I clarify in the footnote but it should be in the body here. Also, “score” is not of interest. We don’t use it in the experiments, don’t waste attention on it.</i>	104
<i>[C5] todo: J: this chapter is actually unfinished. (I can’t believe I missed this. OK. There’s a little more to say on the dataset itself; set yourself up in particular for the parts of the data that are important in the experiments. Make note of the changes between S1 and S2; I think I’d meant to look this up, forgot, and thought the final section was complete, which is a serious oversight, but there’s not too much left to write here.)</i>	104
<i>[C6] todo: My main concern with this chapter at time of writing (10-10-23) is that it goes into too much trivial detail (e.g. types passed into different functions), and not enough into substantive detail (experimental design and why things are set up the way they are). Maybe I have enough substantive detail, but just a surplus of trivial detail? Consider this carefully when redrafting and kill your darlings. If you’re not Tom and you’re reviewing this, please be generous in suggesting things to eliminate!</i>	105
<i>[C6] todo: Cite William’s thesis here somehow?</i>	105
<i>[C6] todo: REFERENCE THE RESEARCH QUESTIONS BEING ANSWERED, THEN BRIEFLY SUMMARISE THE EXPERIMENTS BEING CONDUCTED, WHY THEY ANSWER THE QUESTION[S], AND FINALLY GIVE A RUNDOWN OF THE STRUCTURE OF THIS CHAPTER.</i>	105
<i>[C6] todo: There must be tons of good citations for aspects being used to compose together a simulation / model</i>	106
<i>[C6] review: We don’t think this is the case — those citations just don’t exist! Rework to discuss use of AOP in business process modelling, and note that nobody made those models in practice, just tooling.</i>	106
<i>[C6] todo: Add a cross-reference to the chapter on cross-cutting concern simulation accuracy when it exists</i>	106

<i>[C6] todo:</i> maybe cut this list of reasons PyDySoFu is fantastic...	108
<i>[C6] todo:</i> THESE RESEARCH QUESTIONS HAVE BEEN AMENDED SINCE I FIRST WROTE THIS SECTION (I hadn't decided on a wording then) AND THE PLACEHOLDER SUGGESTION THAT PREVIOUSLY EXISTED WAS JUST WHETHER ASPECTUALLY AUGMENTED MODELS COULD BE REALISTIC. THIS CHAP DOES MORE, SO THIS SECTION NEEDS TO BE REWORKED.	109
<i>[C6] todo:</i> Rework to include details on the other RQs, too .	110
<i>[C6] todo:</i> Write a little here on the learning aspects.	110
<i>[C6] todo:</i> Flesh this out as a brief wrap-up of our experimental technique. The end of our "naive" model explanation might be useful to move down here / rework into this para.	110
<i>[C6] todo:</i> Position this figure properly on the page.	112
<i>[C6] todo:</i> Maybe reword? Pretty verbose.	112
<i>[C6] todo:</i> Consider changing fig. 6.1 to only include details about the model of the game itself, leaving the steps about data analysis to the later aspect-applied version (which is the real experiment).	112
<i>[C6] todo:</i> Move fig. 6.1 to .svg for fidelity	112
<i>[C6] todo:</i> These RQs are taken from an early outline of a thesis structure. Maybe they need rewording?	114
<i>[C6] todo:</i> Rearrange either this paragraph of the sections it refers to so they're in order / make sense. At time of writing they refer to 6.3.1, then 6.4, then 6.3.2, then 6.5...!	116
<i>[C6] todo:</i> Investigate: did we ever model confidence with anything other than $c=1$, i.e. a standard logistic curve?	122

[C6] <i>todo</i> : I don't believe I make any note in this section that selecting character pairs based on a known distribution from player data isn't really learning. It's useful, though, because it allows us to demonstrate that using aspects can make us <i>really</i> accurate. So, it answers our research question, kind of, in that the aspectually augmented model is more accurate than the naive one. However, that aspect isn't particularly portable to a new system (we'd still be selecting based on the S1 distribution, not S2, so we'd expect it can't be ported) and doesn't so much change simulated behaviour as patch a part of the model. It's still useful to build models of learning, see which players are accurately represented by different models, and see whether those models are portable. I still have to figure out how to "sell" this properly...	122
[C6] <i>todo</i> : The ordering of this explanation of different uses of aspects is confusing; once the aspects are explained individually, reorder the sections, rework the introduction to the structure of this section, and ensure the outline of the section reflects the changes made. .	123
[C6] <i>todo</i> : Make a reference to the chapter(s) where I explain different seasons of RPGLite. . . .	125
[C6] <i>todo</i> : Do I need code snippets for notes on aspects?	125
[C6] <i>review</i> : This is unfinished! There's an <code>inline todo</code> note commented out in the source below that I've kept to remind me of what I was thinking when I left this in the state it's currently in.	127
[C6] <i>todo</i> : Keeping this in because I might need to explain the aspect, but it's only used in the implementation of choosing based on prior distribution, which I don't believe is ever applied. So! Decide whether to get rid of the writeup of that model. If I <i>am</i> , then get rid of this too. If not, add it to the learning models writeup below and explain it here.	127
[C6] <i>todo</i> : Do we describe an experiment singular or experiments plural in chapter 6?	127
[C6] <i>todo</i> : reference pdsf's explanatory chapter / re-engineering chapter here? unsure which is most appropriate / introduces the concepts first.	128
[C6] <i>todo</i> : Why didn't we make a bayesian model of learning? Should we have? Would this be difficult at all?	130

[C6] <i>todo</i> : Maybe we should have also produced models of other discounting mechanisms, like exponential discounting (which I <i>believe</i> is actually the logically sound way to discount). Would be interesting to see whether some players are well-modelled by other discounting mechanisms. Investigate this!	131
[C6] <i>todo</i> : Find citations explaining hyperbolic discounting	131
[C6] <i>todo</i> : Maybe there's need here for a diagram or some extra equations to make this clearer...?	132
[C6] <i>todo</i> : Some reference to the research question is required here.	132
[C6] <i>todo</i> : I need to write up the strat for using k-fold validation. Should include the parameters selected, evaluation of "best" result, etc.	133
[C6] <i>todo</i> : Discussion required here around the mitigation of randomness — it's required because of RPGLite's inherent RNG — and also explaining how we make use of a grid search in combination with k-fold validation to anneal toward well-fitted parameters of our models of learning for each individual player.	133
[C6] <i>todo</i> : TODO make a diagram of this.	133
[C6] <i>todo</i> : These parameters are required because we run with different parameters for different folds. This subsec should be preceded by another on our strategy for optimmising, so we can introduce k-fold validation and clean up the wordinng in this sec so it makes more sense.	133
[C6] <i>todo</i> : IIRC Tim doesn't love the description list formatting. Probably means I should avoid; consider other ways of laying this out that avoids a big wall of text	133
[C6] <i>todo</i> : This subsection most likely needs to be condensed.	137
[C6] <i>todo</i> : Find citations for each correlation coefficient, and also for justifying choosing between these two specifically (if required).	138
[C6] <i>todo</i> : I'm not sure I've defined how I represent (or calculate) the ordered list of preferences yet. If not, that should live somewhere, I suppose.	139
[C6] <i>todo</i> : A little clumsy — reword this para? Is it suffient?	140

<i>[C6] review:</i> At time of writing metagame is not defined, but this should be introduced in chapter 5; there's a corresponding revision note there.	140
<i>[C6] todo:</i> TODO note that players are removed, and then replaced, so that the playerbase is of constant size but varying experience! Replacement happens immediately.	141
<i>[C6] todo:</i> Mention boredom period controlling how often we check that players are bored. . . .	141
<i>[C6] todo:</i> TODO note that boredom can be disabled	141
<i>[C6] review:</i> I end list items here with semicolons. I should be consistent through the thesis, and decide whether semicolons are appropriate.	141
<i>[C6] todo:</i> A note here that we're using AOP, but don't actually target cross-cutting concerns in our models of learning, which is interesting.	142
<i>[C6] todo:</i> To be honest, this might be the best explanation of the experimental design I've produced so far. Maybe I should be re-using this in the experimental design section?! . . .	142
<i>[C6] todo:</i> A discussion of how datasets are evaluated. We produce <i>many</i> datasets which have annealed to the highest stat and P val we could find, but they're split across many folds. How do we determine whether a set of parameters works "enough"? We find the highest correlation val and lowest pval which work across a majority of folds, then test against the complete dataset to be absolutely certain that correlation doesn't drop. This gives us the best parameters for a player: the parameters were reliably annealed to across many folds to give statistically significant correlation, retained their correlation against a testing set, and further retained their correlation against the entire dataset.	143
<i>[C6] todo:</i> Is there a pattern in players getting reliable correlation and having played many games? Maybe we just don't have enough games to do this for most players. Check this; could be worth a note toward the end, in a discussion or similar.	144
<i>[C6] todo:</i> This is an explanation of the analysis script, not the process of annealing; we should already have that in the experimental design section (currently 6.5, at time of writing). . .	144

<i>[C6] todo:</i> TODO: add notes explaining what an acceptable pval and correlation stat are. This will be useful to refer back to in the next subsection, where not all players are simulated by each model.	144
<i>[C6] todo:</i> Explain the results & relate them to the final RQ.	145
<i>[C6] todo:</i> I suspect I'll move this to the discussion of the next experiment, so that the aspects using the known distribution are presented as an investigation of the first research question, and the aspects using the confidence-based model of learning are presented as an investigation of the second.	145
<i>[C6] todo:</i> I'm increasingly realising that the aspect drawing from a known distribution is both more accurate and more portable than our models of learning: we can apply it to S2 and just draw from season-specific data to get the correct results. It's actually way better than our models of learning. However, the models of learning let us investigate specific players' behaviour, so they let us use the naive model to investigate the bahviour of real-world players, which is still quite valuable, but might want a different approach to the write-up. I'm pretty lost with this chapter — need Tim's input, I think.	145
<i>[C8] todo:</i> is suitable for simulation purposes? It's about showing that we <i>can</i> use aspect orientation appropriately in simulation environments, and that aspect orientation can also lead us to realistic and nuanced simulations, too. Go back through the pdsf and lit review chapters to make this argument properly.	151
<i>[C8] todo:</i> Add a RQ at the end of each research opportunity; this can and should be a contribution of the thesis, and the RQs would help to sell it as such.	151
<i>[C8] todo:</i> Find citations for World3 model	152
<i>[C8] review:</i> "Freak event" might be too informal here but I'm not sure what to replace it with.	153
<i>[C8] review:</i> If I've got time, I'd love to push my codebase augmenting a World3 model to achieve just this with predictive implications for the 2008 financial crisis and the covid pandemic.	154

<i>[C8] todo:</i> Cite popper, any point including studies that do this too? It's fairly table stakes as a concept...	155
<i>[C8] review:</i> Maybe too informal, but I don't know, maybe live a little...?	155
<i>[C8] todo:</i> find a citation for modern algorithmic / robotic drug development	155
<i>[C8] review:</i> This subsec is already quite long, and including this point makes it significantly longer. Maybe that's not a prob; maybe this should be a separate point / subsubsec; maybe it can live in a footnote; maybe it should be omitted completely.	155
<i>[C8] review:</i> I've made this an enumerated list; is it better as a paragraph?	157
<i>[C8] review:</i> "obvious" is a little too informal, not sure what to replace it with here.	159
<i>[C8] review:</i> This might be a little rambly — consider omitting, rewording, or factoring into a footnote.	159
<i>[C8] todo:</i> We can get more precise than chapter references; find the sections / subsections.	159
<i>[C8] review:</i> Are there any other reasons as to why models might be incorrect (and instrumentation could identify this by boundary checking)?	160
<i>[C8] review:</i> Is this point worthwhile? I cited it in IIRC an experimental chapter, but I'm unsure it's worth including. For future Tom to revisit.	162
<i>[C8] review:</i> Consider refactoring this into a subsection of that sec.	164

<p><i>[C8] review:</i> Possibly too grandiose? I don't think so really, but I wouldn't want to present it as being more than it is, and worry I oversell here. Maybe some references to literature discussing the expected scope of PhD theses (which exists! I have it in KeepIt, Tim sent me some early in the PhD) and an explanation that it's not outside the realm of a normal thesis to make contributions with theoretical implications — see REST thesis or computational trust — it does warrant a beefy future work chapter and an exploration of the research this enables (because it <i>does</i> enable lots of future research) is a contribution in its own right. Which isn't necessarily grandiose, but can be oversold as grandiose. Maybe comparisons to REST don't help that case, actually...</p>	166
<p><i>[C8] review:</i> Maybe go into more detail on what was mentioned in specific cross-referenced sections here? A long list of sections this chapter contains without any extra explanation isn't very satisfying, even if it's useful within this framing of exploring what Kuhn meant and how it relates to this thesis.</p>	168
<p><i>[C9] todo:</i> Duplicate citation to clean up</p>	180

Chapter 1

Introduction

Consider a research software engineer, responsible for the development and maintenance of a well-adopted model of some socio-technical system — perhaps a public transport system, the spread of a disease in a pandemic, or a software development team’s day-to-day work. As the model’s use spreads and users’ requirements broaden, how might our engineer manage their maintenance burden? What tools and techniques might they make use of to ease their task? Their tool might have originally modelled the interactions of individual developers in a team (to extend the latter example), and offered parameters for development methodologies or team sizes; subsequent researchers might look to run simulations of the team with biased behaviour such as illness or tiredness, or different experience levels in the team, or communication breakdowns between individuals, or the impact of a change in management or project direction. What then?

Typically, if a required change would alter the model significantly, researchers looking to adopt the tool for new purposes would fork it (assuming the source is available to them under a permissive enough license) and would make whichever modifications suited their needs. Other researchers might make their own forks. If the change was small, or if the new behaviour was enabled in configuration, then it might be merged into the original tool. The contribution is thus disseminated to the research community who adopted the original tool without the need to migrate to a fork. However, if a future team wanted to research a combination of factors — say, simulating the impact of a change of management on teams with different levels of experience and communication quality — they would repeat

the process, producing another fork with their own implementation of this particular combination. As modifications are built on top of each other, the logic for each possible behaviour is interwoven with that of the original model.

Such a codebase would be increasingly difficult to maintain. For example, the abstractions used to modularise it may have elegantly separated different concerns into different modules at first, but abstractions are often domain-specific; as the model becomes increasingly general-purpose, its abstractions used to separate concerns accumulate technical debt. In addition, the behaviours added by different teams may not make sense when enabled at the same time, such as a change representing remote work and a change representing the spread of a virus in different office settings; the tool's source code risks becoming confusing and unwieldy.

1.1 A primer on aspect orientation

Explain aspect orientation briefly. Some useful notes for the explanation:

- AO originated in xerox parc, first described in [Kic+97]. There are lots of weaving mechanisms for regular, static aspect orientation, and there's a good early survey of them all (and implementation in a custom OO language specifically for this!) in [MK03].
- AO has some forebears: metaobject protocols, subject-oriented programming, adaptive programming, composition filters. The latter three are described by [CBB19] as being alternative *kinds* of AO — I disagree, but they're certainly attempting similar things.
- The original and still most widely used AO implementation is AspectJ, which comes with its own aspect language. It's grown over the years and is used sometimes in industry [citation needed...]. A smaller alternative would be Spring AOP **find a citation for spring AOP**

1.2 Prior Work

Write a section for the introduction describing the work done on pdsf before this, to delineate

where we're starting from and avoid any claims of plagiarism. This can be short, the first sec of the lit review is a proper discussion, but the tool should be mentioned here. See ?? for what already exists.

1.3 Terms & definitions

Complete the glossary in section 1.3.

Decide whether terms like BPMN, simulation & modelling, etc also belong in the glossarysection 1.3.

Dejice's thesis has a similar glossary; how did he structure it? Wasn't an appendix. Was it a section of the intro or something else?

Aspect-Oriented Programming

Cross-Cutting Concern

Aspect

Advice

Joinpoint

Pointcut

Weaving

AspectJ The original aspect orientation framework, with language extensions to describe pointcuts and aspects.

Target The procedure an aspect is applied to via a join point, to affect advice.

PyDySoFu

Chapter 2

Relevant Literature

This thesis presents an aspect-oriented approach to simulation and experimental design, tooling to support these endeavours, some empirical assessment of both the paradigm and its application in the domain of simulation and modelling. In particular, the presented research makes use of aspects to model hypotheses and the complexities of variable and unpredictable behaviour in the simulation of socio-technical systems. Their evaluation measures code quality improvements, which are commonly claimed in aspect orientation research . systems.

Relevant literature in this topic typically comes from a variety of fields which do not overlap significantly, meaning that this review must cover segments of several partly related fields. The presented material is eclectic in nature as a result. For this reason, the literature reviewed in this chapter includes present context for unfamiliar readers as well as a motivating case for the work which follows in later chapters.

philosophy toward software engineering and modularity, and some criticisms of its high-level design. Other sections, in particular section 2.1 and section 2.2, expand on this introduction by investigating implementations, technologies and use cases of particular interest.

add citations
where code
quality im-
provements
are claimed
in AOP re-
search

There's a
paragraph
missing here
explaining
which chap-
ters follow,
why, and in
what order.
I'm omitting
it at the mo-
ment because
I suspect I'm
about to re-
order them.

2.0.1 Motivations & Philosophy Underlying Aspect Orientation

Modularity is considered a key trait of maintainable, flexible, and legible programs [parnas_1972]. Modern modular design techniques are often concerned with segmenting a program's logic or data into separate units. The standard approach to designing with modularity in mind in many industrially-relevant languages today is object-oriented programming, where data structures and selected program logic are modularised as classes. Cite some OO languages, papers noting their industrial use / the "success" of object orientation.

Some concepts in software engineering are not easily modularised by techniques such as aspect orientation: guarding against unsafe concurrency usage, manual memory management, and logging are all examples of program components which are not readily modularised using an object-oriented or similar approach: the logic of these "cross-cutting concerns" is common to all modules. Programmers looking to separate cross-cutting concerns into separate modules look to address two problems [Kic+97]:

- ① *"Tangling"*, where program logic essential for a program's intended purpose is intermixed with ancillary code addressing cross-cutting concerns, thereby making essential logic more difficult to maintain;
- ② *"Scattering"*, where program logic for cross-cutting concerns is strewn throughout a codebase, making maintenance of this code more difficult.

The existence of cross-cutting concerns is therefore expected to make maintenance of both ancillary program logic and a program's core logic more difficult. Addressing this, Kiczales et al. introduced the notion of aspect-oriented programming [Kic+97]. Aspects are simple to define through their component parts:

- A *"join point"* defines some point in a program's execution (usually the moment of invocation or return of some function or method).
- *"Advice"* defines some behaviour such as logging, which which can conceptually happen anywhere in program execution (i.e. what's defined would typically represent behaviour which cuts across many parts of a codebase).

- An “*aspect*” is constructed by composing this advice with “*point cuts*”: sets of join points that define all moments in program execution where associated advice is intended to be invoked.
- An “*aspect weaver*” then adds the functionality defined by each aspect by adding the functionality defined by its advice at each join point defined by its point cut.

The definition of join points and advice, or how weaving occurs, is a matter left for aspect orientation frameworks and languages to define. In employing the technique, aspect oriented programming aims to separate cross-cutting concerns into aspects, removing the aforementioned repetitive code from the logic implementing a program’s functional behaviour so that additional pieces of functionality — logging, authentication, and so on — can be maintained in only one place in a codebase (thereby simplifying their maintenance and comprehension), and remaining program logic can be understood and maintained without the overhead imposed by the previously tangled cross-cutting concerns.

In [Kic+97], Kiczales et al. see these engineering concepts as universal throughout business logic, motivating the aspect-oriented approach for the first time. The authors present an implementation of AOP in Lisp, and compare implementations by way of e.g. SLOC count in an emitted C program to a comparable, non-AOP implementation, with two examples (its use in image processing and document processing). They find the idea — which they note is “young” and note many areas where research might help it to grow — can successfully separate systemic implementation concerns such as memory management in a way that reduces program bloat and simplifies implementation. It is noted that measuring the benefits of their approach quantitatively is challenging.

Tooling followed the theoretical work presented by Kiczales et al. [Kic+97] with a demonstration and subsequent technical description of AspectJ, a Java extension for aspect oriented programming [Hil+00; Kic+01]. AspectJ was introduced to satisfy the research community’s need for a tool with which to demonstrate the aspect-oriented paradigm in case studies. The tool is intended to serve as “the basis of an empirical assessment of aspect-oriented programming” [Kic+01]. The library makes use of standard aspect-oriented concepts: Pointcuts, Join Points, and Advice, bundled together in Aspects. They define “dynamic” and “static” cross-cutting, by which they refer to join points at specific points in the execution of a program, and join points describing specific types whose functionality is to be altered in some way. Their paper describes only “dynamic” cross-cutting, but presents tooling support,

architectural detail of its implementation, and the representation & definition of pointcuts in AspectJ. AspectJ is compared to other tools for aspect orientation and related decomposition-focused paradigms, and the authors are explicit about their approach being distinct from metaprogramming in, for example, Smalltalk or Clojure.

Filman, Friedman, and Norvig isolate properties of aspect orientation which they assert are definitive of the paradigm [FFN00]. Specifically, they claim that aspect oriented programming should be considered in terms of “quantification” and “obliviousness”:

“AOP can be understood as the desire to make quantified statements about the behaviour of programs, and to have these quantifications hold over programs written by oblivious programmers. [...] We want to be able to say, “This code realises his concern. Execute it whenever these circumstances hold.”

These concepts became core concepts in aspect orientation literature alongside “tangling” and “scattering”. Filman, Friedman, and Norvig give no concrete definition of the terms in their original paper and cite no sources for the terms; for the purposes of this thesis, we therefore provide the following definitions of the terms:

Quantification is the property of specifying specific points in a program in which that program should change;

Obliviousness is the property of a codebase that it contains no lexical or conceptual reference to advice which might be applied to it, and of the programmer of a target program that their code may be amended by an aspect programmer.

Filman, Friedman, and Norvig write about aspect orientation “*qua programming language*”, and so theorise around aspect orientation as a paradigm independent of a particular application. They are therefore able to arrive at conclusions about the paradigm in the abstract, and can identify concerns for future investigation for researchers in the field and design goals for developers of aspect orientation tooling. They note:

“Better AOP systems are more oblivious. They minimise the degree to which programmers (particularly the programmers of the primary functionality) have to change their behaviour to realise the benefits of AOP. It’s a really nice bumper sticker to be able to say, “Just program like you always do, and we’ll be able to add the aspects later.” (And change your mind downstream about your policies, and we’ll painlessly transform your code for that, too.) [FFN00]”

Here, Filman, Friedman, and Norvig theorise about the advantages they expect aspect orientation to bring software engineers. However, their writing follows the original from Kiczales et al. [Kic+97] by only three years, and in the following decades no significant literature or case study seems to evidence their expectations [Ste06]. Criticisms of aspect orientation’s effectiveness are discussed further in section 2.0.2.

Aside from evidence of practical benefits of aspect orientation, claims made by Filman, Friedman, and Norvig such as the “painless transformation” of code when a developer changes their mind during the development process — a theorised benefit of obliviousness — is incompatible with earlier writing on modularity. Yourdon and Constantine [YC79] assert:

“The more that we must know of module B in order to understand module A, the more closely connected A is to B. The fact that we must know something about another module is a priori evidence of some degree of interconnection even if the form of the interconnection is not known.”

Aspect orientation’s critics describe similar incompatibilities with existing best-practices [Prz10; CSS04], as well as the lack of empirical evidence for the benefits of obliviousness [Ste06]. Claims about “better” aspect-oriented systems being more oblivious should therefore be regarded as *suggestions* from the literature. While obliviousness and quantification are useful concepts in discussing research in the field, they also give context for the research community’s perspective that obliviousness and quantification are design goals for aspect oriented programmers [AspectCplusplusDesignImp; Kel08; CM06] (though Leavens and Clifton [LC07] suggest they may be best applied in moderation).

Feels like the full literature on AOP is missing here. What is the

2.0.2 Criticisms of Aspect Orientation

The growing aspect-oriented programming research community collected both proponents and detractors of the paradigm. The opposition to aspect orientation is of particular relevance to this thesis in two regards. First, this literature review discusses advancements in the field relevant to the work presented in this thesis. This should be understood within the context of some perceived weaknesses of aspect orientation, which helps to frame an understanding of the literature reviewed. Second, the following chapters introduce contributions which address some criticisms of aspect-oriented programming, meaning that the criticisms of the paradigm writ large — and the properties of work published in awareness of those weaknesses — motivates some research presented in later chapters.

Constantinides, Skotiniotis, and Stoerzer published an early critique of aspect-oriented programming [CSS04] which notes similarities between the paradigm’s core concepts and GO–TO statements. Constantinides, Skotiniotis, and Stoerzer uses the comparison to demonstrate fundamental issues with the paradigm’s philosophy: use of GO–TO statements is widely accepted to be bad practice [Dij68], and their infamy has become an in-joke for academics and language enthusiasts [Cla73]. They note that GO–TO statements disorientate a programmer by way of “destroying their coordinate system” — referring to a developer’s technique for navigating and understanding code — which results in uncertainty about both a program’s flow of execution and the state of a program at different points of its flow. Constantinides, Skotiniotis, and Stoerzer uses the comparison to GO–TO statements to question whether aspect-oriented programs can have a consistent coordinate system for developers: GO–TO statements lack obliviousness as they are visible in disrupted code, whereas aspects are not represented structurally within a program due to their oblivious design. This complicates a developer’s understanding of where and how flow is interrupted. They draw comparison between aspects and COME–FROM statements, an April Fools’ joke where a claimed improvement over GO–TO is developed by removing the latter’s structural component [Cla73], and conclude that existing engineering techniques provide similar benefits without a trade-off in program legibility. In particular, Dynamic Dispatch is identified as a preferred alternative.

Steimann makes a similar but more thorough critique of aspect-oriented programming [Ste06]. They express concern that the popularity of aspect-oriented programming — which was nearly 10 years

old at time of publication — was founded on the perception that it would solve real-world engineering problems, yet no proof existed that it was effective in practice. Steimann notes that most papers are theoretical in their discussion on tooling, that examples were typically repetitive, and that the community's discussion was concerned more with what aspect orientation could be used for than whether it worked in practice. They present a comparison between aspect orientation and object-orientation, where aspect orientation's claimed properties and principles are examined in detail, and the impact on software engineering is reasoned about from a skeptical perspective. They compare the paradigms' specific claims that they both support improved modularity against classic papers on the subjects.¹ Steimann presents a philosophical examination of aspect orientation and assesses the paradigm against its purported merits, discussing whether we should accept the claims made by the aspect-oriented programming community. Aspect orientation's promise of unprecedented modularisation is presented as unfulfilled, and Steimann reflects critically on the state of aspect orientation research at the time. However, they conclude by proposing that the aspect-oriented approach could have a legitimate alternative use-case — other than as a *general-purpose* technique for modularising cross-cutting concerns — where it could be shown to be effective empirically.

Similar sentiments are shared by Przybylek [Prz10], who examines aspect-oriented programming in the context of language designers' quest to achieve maintainable modularity in system design. They frame the design goals of aspect orientation as being to represent issues that “cannot be represented as first-class entities in the adopted language”. Przybylek questions whether the modularity offered by aspect orientation can really be said to make code more modular. In particular, they distinguish between lexical separation of concerns and the separation of concerns originally discussed by Dijkstra [Dij82]. They assess principles of modularity — modular reasoning, interface design, and a decrease in coupling — and find that the aspect-oriented paradigm can detrimentally impact the expected benefits of proper modularisation in a program . They suggest that aspect orientation's claimed benefits are a myth repeated often enough to be believed.

¹In particular, they compare against literature by `parnas_1972empty citation`.

Why? Presumably because they use the wrong type of modularity (Lexical?) — if so, make this clear here!

This review is arguably

2.0.3 Alternatives to Aspect Orientation

Aspect-oriented programming's goal of modularising cross-cutting concerns is shared by other paradigms. As discussed in section 2.0.1, the work first introducing aspect orientation by Kiczales et al. [Kic+97] makes note of similarities to reflection, metaprogramming & program transformation, and subject-oriented programming. They also observe that other disciplines have introduced "aspectual decomposition" independently.

this needs more explanation. What were Kiczales et al. trying to do, what was the context, what did they demonstrate?

System Diagrams

The example of pre-existing aspectual decomposition by way of diagramming given by Kiczales et al. [Kic+97] is in physical engineering. To give a concrete example from their description, differing types of diagrams when engineering a system such as thermal and electric diagrams of a heater are described as "aspectual" because of the modular nature of the diagrams; though there might be many diagrams of different kinds, they compose together to give an overview of the system being designed.

Reply/Q: I covered this a fair bit in the earlier parts of this section. Maybe this was unclear? This bit is already covered, so fixing earlier notes might be better than adding notes here. I've since restructured & reframed, hopefully enough to make this clearer, but I'm curi

Similar diagramming techniques have independently arisen in other domains since. The Obashi dataflow modelling methodology[WC10] by Wallis and Cloughley models all possible paths of dataflow through "B&IT" (business and IT) diagrams, where business-specific concerns (people, locations, groups, and business processes such as payroll, stock-check or budgeting) are modelled alongside IT concerns such as applications supporting business processes and the software and hardware infrastructure supporting them. Modelling dataflows in this way allows for a comprehensive understanding of assets and business processes. However, in order to understand how data flows between specific assets within a B&IT, sub-graphs ("DAVs", or Dataflow Analysis Views) denote specific pathways through which data flows between source and sink assets. Alternatively, a B&IT can be viewed as a composition of all possible DAVs within an organisation. Dataflows are therefore broken into different diagramming techniques and specific business concerns can be described independently of others, even if these concerns interact in their dataflow pathways (and, therefore, cutting across each other). Obashi therefore allows for the aspectual decomposition of business processes, through the description of an organisation by individual dataflow analysis views, which compose into an overall model of a system in a B&IT diagram. Obashi models are an instance of aspect orientation which were designed

for simplicity and comprehension[WC10; Seo11], but trade this for domain-specificity.

Add notes on Bigraphs here

Add notes on Holons here, maybe

Metaprogramming

Metaprogramming is identified as a precursor to aspect orientation by the original paper on the paradigm [Kic+97]. Research into the use of metaprogramming to simplify the composition of software modules was undertaken at a similar time by Keller and Hölzle [KH98], who introduce a technique they name “Binary Component Adaptation” (BCA). Their research into BCA the difficulties involved in the integration of software components, particularly considering their evolution over time where components are re-used with differing requirements. By modifying binaries directly, incompatibilities between a program and one of that program’s dependencies can be resolved by way of mutating either after compilation. Their implementation defines a representation for the modification of pre-existing Java class binaries, the output of which can be verified as also being valid Java class binaries. Keller and Hölzle claim that BCA allows for dynamic modification of programs with little overhead. They believe BCA is unique in its combination of features, which include engineering concerns such as typechecking code which is subject to adaptation and its obliviousness to source implementation, as well as guarantees that modifications are valid even for later iterations of the program subject to adaptation.²

Metaobject protocols describe the properties of an object’s class (including, for example, its position within a class hierarchy) in an adaptable manner [KDB91]. Espák [Esp04] note that this technique can be used to implement aspect orientation, therefore providing at a minimum the same functionality, though they achieve this through reflective programming techniques and are designed with metaprogramming as a primary goal as opposed to modularisation of cross-cutting concerns [KDB91; Sul01].

²BCA shares concepts with aspect orientation, but is also a promising technology for the introduction of process variance; see section 2.3.2 for a discussion.

Wouldn't the paragraph below fit better in the section on program mutation?

Reply/Q: I'm not sure exactly what section this is, but have reworded the below to fit better within this section (BCA is a relative of AOP) and added a cref to where I think it might be added. Unsure though. To discuss and fix the cref. I haven't added notes on BCA anywhere else. Or, was it intended to

Engineering Techniques with Related Aims

Multiple-dispatch, where methods on objects are chosen to be run based on the properties of the parameters passed at point of invocation, allows for oblivious decomposition without the need for a weaver [DGM08], although this does not support the goals of aspect orientation in totality. For example, a programmer might want their program to exhibit differing behaviour when methods are called with differently-typed arguments, which is supported by multiple dispatch. However, they might instead want their program to exhibit some additional behaviour whenever a method is invoked, such as logging, but might not want to implement logging alongside the rest of their method implementation for clarity or length reasons. Multiple dispatch therefore offers comparable but different functionality to a software engineer.

Engineering patterns such as decorators provide similar functionality to aspects [FBS17], in that cross-cutting concerns can be separated into their own module, but they differ in their approach to obliviousness: decorators annotate areas of a codebase they are applied to, and therefore do not offer obliviousness as aspects do. Decorators allow for additional functionality to be written as a separate module and applied as a wrapper around a function definition. **A diagram or Python example might help here, as per Tim's suggestion below.** A function with a wrapped definition is replaced by a function returned by the decorator, which takes the original definition as an argument [Gam+95]. Additional logic can therefore be simply applied before or after a function. While this achieves a similar effect to aspect orientation — as logic is added before and/or after the original logic of a function — its design principles are different as annotating the function's definition directly marks it as altered, and

so the original definition cannot be oblivious to the change.

2.0.4 Subject-oriented Programming

Subject-oriented programming is a paradigm which **Note: there's an argument that SOP has seen more success in practice than AOP: Golang's interfaces are subjective views on a struct, and I believe Rust's Traits are too — confirm this! — by contrast, there's some limited use of AOP in the wild, but related design patterns have arguably supplanted them for modularising cross-cutting concerns (I'm thinking of Decorators). I'm not aware of any literature which backs this up.**

Would some illustrative examples of these different approaches help?

Reply/Q: Yes, I'm sure they'd help! Diagrams to add here.

Incomplete! Add subject-oriented pro-

relevant papers:

- Harold Ossher, Peri Tarr, William Harrison, Stanley Sutton, N Degrees of Separation: Multi-Dimensional Separation of Concerns, Proceedings of 1999 International Conference on Software Engineering
- Subject-oriented programming: a critique of pure objects
- Whatever originally defined SOP
- Founding papers for Hyper/J and/or Hyper/C

2.1 Dynamic Aspect Weaving

2.1.1 Approaches to Dynamic Weaving

The performance issues noted by Popovici, Alonso, and Gross are explored in more detail by Chitchyan and Sommerville [CS04], who present a review of early dynamic aspect orientation techniques. They compare AspectWerkz, JBoss, Prose, and Nanning Aspects through the lens of the authors' prior work on dynamic reconfiguration of software systems. By comparing different implementations of dynamic weaving, Chitchyan and Sommerville contribute a categorisation of the tools' approaches:

- ① "Total hook" weaving, where aspect hooks are woven at all possible points;
- ② "Actual hook" weaving, where aspect hooks are woven where required;
- ③ "Collective" weaving, where aspects are woven directly into the executed code, "collecting the aspects and base in one unit".

As Chitchyan and Sommerville focus on software reconfiguration rather than the mechanics and design of dynamic aspect weaving, their analysis of the reviewed tools is of less relevance to the work presented in this thesis than their generalisation of dynamic weaving. However, their review discusses

I think this comes before the critique section perhaps? Need a general introduction to AOP techniques and techs, including this?

T: Have I actually explained what a dynamic weaver is? Think not.

T: maybe this is a suitable introduction to dynamic weaving, if I add an explanatory paragraph at the beginning?

the trade-offs of the three approaches identified. Chitchyan and Sommerville propose that total hook weaving allows flexibility in the evolution of a software product, at the expense of the performance of that product; this contrasts collected weaving, which shifts overhead out of the codebase and into the maintenance effort. Actual hook weaving is positioned as a compromise between the two, offering the best approach for none of their criteria but never compromising so much as to offer the worst, either. This suggests merit in a tool designed to flexibly offer any weaving approach appropriate for the task at hand.

Chitchyan and Sommerville note that one could use many of the systems they describe in practice. Though the paper is an early publication in the field, no tool the authors review offers all three dynamic weaving approaches, and none offers collective weaving alongside either kind of hook weaving.

AspectJ belongs here.

2.1.2 PROSE

One implementation of dynamic weaving is PROSE [PGA02; PAG03], a library which achieves dynamic weaving by use of a Just-In-Time compiler for Java. Popovici, Gross, and Alonso motivate the project by identifying aspect orientation as a possible solution to software's increasing need for adaptivity. Mobile devices, for example, could enable a required feature by applying an aspect as a kind of "hotfix", thereby adapting over time to a user's needs. Other uses of dynamic aspect orientation they identify are in the process of software development: as aspects are applied to a compiled, live product, the join points being used can be inspected by a developer to see whether the correct pointcut is used. If not, a developer could use dynamic weaving to remove a misapplied aspect, rewrite the pointcut, and weave again without recompiling and relaunching their project.

The conclusion Popovici, Alonso, and Gross [PAG03] provide indicates that some performance issues may prevent dynamic aspect orientation from being useful in production software, but that it presented opportunities in a prototyping or debugging context. The PROSE project explores dynamic weaving as it could apply in a development context, but the authors do not appear to have investigated dynamic weaving as it could apply to simulation contexts, or others where aspect-oriented software does not constitute a product.

Their observation that aspect-oriented programming could be used for the purpose of adaptation

or software prototyping instead of modularisation is an example of an alternative use case for aspect orientation suggested in Steimann’s critique [Ste06], as discussed in section 2.0.2.

2.1.3 Nu

et al. propose a new aspect-oriented invocation mechanism, which they call “Bind” [+06]. Bind’s design is motivated by opportunities to improve modularity from a design perspective: et al. assert that “scattering” and “tangling” can be introduced into a codebase after weaving with some weaver implementations. This complicates the use of compiled aspect-oriented code, and the development and execution of unit tests on such a codebase. Bind alleviates this issue by allowing a developer to choose when aspects are applied. This new mechanism is presented as an alternative to the weaving of aspect hooks for load-time registration into target code in the style of PROSE [PGA02; PAG03] and the direct weaving of aspect invocations in the style of AspectJ [Kic+01].

In order to demonstrate Bind’s approach to simplifying post-weave codebases, **rajan2006_towardsao_invocation** also present “Nu”, an aspect orientation framework written in .NET supporting Bind. Nu’s design is explained and its implementation presented, which are designed to promote granularity in join point specification. What results is a flexible model for aspect orientation which is demonstrated to satisfactorily emulate many other paradigms and tools, such as aspect orientation in AspectJ, subject-orientation in HyperJ, and Adaptive Programming.

Dyer and Rajan explain in more depth than in the design and implementation of the Bind mechanism and the implementation of the Nu framework [DR10]. A more technical discussion is presented, in particular on implementation details including optimisation and benchmarking, largely against AspectJ. Notably, the implementation discussed is a Java implementation, rather than the .Net implementation presented in [rajan2006nu]. Many aspect orientation frameworks are language-specific; the existence of Nu’s implementation on multiple platforms highlights the work’s main contribution in the design of the Bind primitive, rather than the framework itself.

It is noted that it is “very common in aspect-oriented programming research literature to provide language extensions to support new properties of aspect-like constructs”. **rajan2006_towardsao_invocation** look to provide similar extensions to virtual machines, noting that their mechanism is a suitable can-

I don’t define adaptive programming. Maybe I should have it as an alternative? End of this section, after subject-oriented programming?

didate to introduce aspect orientation directly in a language's virtual machine. They position the project as a general model of aspect-oriented programming which can flexibly represent a variety of existing approaches. Bind fulfils this aim by providing a single mechanism which can be used to achieve many weaving techniques; however, case studies demonstrating that Bind can be used to make aspect orientation more practically effective for software developers does not appear to have been published.

2.1.4 Dynamic Weaving in Embedded Systems

Gilani and Spinczyk observe that while there are different approaches to dynamically weaving aspects, no approach is suitable for an embedded environment [GS04]. This is due to these systems' low power and limited memory. Gilani and Spinczyk propose a framework for these situations through which weavers can be assessed for their suitability in a given environment, or generated from a set of desired features.

Gilani and Spinczyk define families of weavers by grouping the the environments they can be suitably applied to, separating them in particular by their trade-off between dynamism and resource use. There is an implication that dynamism and resource use are broadly proportional, presumably because even a carefully crafted "actual hook weaver" or JIT-compiled "collective weaver" in the parlance of Chitchyan and Sommerville [CS04] carries runtime overhead by virtue of the weaving mechanism used. A "collective weaver" embeds aspect invocations directly into their join points, which eliminates the need for an intermediary hook invocation or additional compilation. The embedded systems of interest to Gilani and Spinczyk have memory in the range of ~30kb, where these overheads could represent significant resource use.

Aspect oriented programming's criticism can often be that it doesn't know what it "aims to be good for" , and the dynamic weaving of aspects may not be "good for" resource-constrained environments at all. As Gilani and Spinczyk identifies a trade-off between dynamism and resource economy, a highly constrained environment would not be able to take advantage of dynamic weaving by definition. Additionally, because the anticipated benefits of aspect-oriented programs are not observed in practice, the paradigm may be poorly matched to any use-case where pragmatism is a concern. From the

Where's
this quote
from? Cita-
tion needed.
Think it
could be
Przybylek
[Prz10] or
Steimann
[Ste06].

framework proposed by Gilani and Spinczyk [GS04], we might observe that researchers should seek other contexts in which to apply aspect oriented programming. Support for this observation can be found in criticisms of aspect-oriented programming from Steimann [Ste06], whose position that alternative use-cases of aspect orientation should be explored is discussed in section 2.0.2. Examples of alternative uses include Popovici, Gross, and Alonso’s proposal of using aspect orientation for software prototyping, and the proposal presented by Gulyás and Kozsik [GK99] that simulation & modelling is a more appropriate field. Both of these proposed uses of aspect orientation involve a different resource economy than that imposed by embedded systems. A discussion of aspect orientation’s use in simulation & modelling can be found in section 2.2.

2.2 Aspect Orientation in Simulation & Modelling

This thesis is concerned with the use of aspect orientation in simulation & modelling codebases; it is therefore necessary to review related research in simulation and modelling. This section reviews literature contributing aspect-oriented tooling and the philosophy of aspect-oriented experiment design in the simulation & modelling community.

2.2.1 Suitability of Aspect Orientation in Experimental Codebases

Gulyás and Kozsik [GK99] observed that, in the study of complex systems through software models, the codebase produced serves two purposes: the experimental subject, and the observational apparatus used to conduct the experiment itself. They use this framing to identify that the observation of a program’s state constitutes a cross-cutting concern. In order to reduce scattering and tangling in experimental codebases, Gulyás and Kozsik theorise that aspect-oriented programming may separate the logic of observation from the core logic of a simulation or model.

The approach is demonstrated via their Multi-Agent Modelling Language (MAML). MAML was designed to enable aspect-oriented simulation of agent-based models and was implemented using the SWARM simulation system [Hie+94], a domain-agnostic framework for agent-based simulation. While SWARM takes the form of a collection of C libraries, MAML is implemented as a domain-specific

modelling language, allowing it to support aspect-oriented programming as a language feature.

MAML's aspect orientation effectively makes use of Observer patterns to measure a simulation's state. This enables a researcher to observe an experiment without the necessary logic being scattered or tangled in their domain model. Gulyás and Kozsik find that aspect-oriented programming provides an intuitive and straightforward method by which simulated experimental systems can be composed. They note that MAML's simplicity and its philosophy on modelling are more "satisfactory" than Swarm's standard approach. The team report that MAML's implementation was more complex than initially conceived: the patch unix tool was intended for use as their weaver, though the team eventually developed a transpiler from MAML to Swarm instead (which they name `xmc`). The deciding factors for the development of a custom transpiler are not discussed.

In addition to presenting tooling for aspect oriented simulation, Gulyás and Kozsik [GK99] theorise about the potential benefit of applying aspect orientation to simulation & modelling. They observe that there may be benefits beyond improvements to modularity and a reduction of tangling & scattering. In particular, their work discusses specific scenarios in which the *type* of separation of concerns offered by aspect orientation is desirable, and the engineering approach to achieving the aim reasonable.

T: If there's time when redrafting, note which scenarios they distinguish between specifically.

This distinguishes the work in comparison to most other research on aspect orientation. Many papers describe the expected benefits by simply drawing from existing literature and the claims made in Kiczales et al. [Kic+97]'s first paper on the subject.

it seems like this might be an important exposition, so perhaps more details?

This is a rare example of philosophy on aspect oriented programming's suitability in a particular use-case. Gulyás and Kozsik's work is of particular importance in this review, because the domain they identify as particularly suitable is simulation & modelling, which is the subject of this thesis. That aspect orientation might be well suited to separating observer and experiment partially motivates research in later chapters, which investigates the use of aspects to *modify* simulated behaviour rather than simply observing it. Further discussion follows in section 2.4.

Reply/Q: This is covered in the criticisms of AOP. Maybe I should just make a reference here,

2.2.2 Aspect Orientation in Discrete Event Simulation

Aspect-Oriented Implementations of Simulation Frameworks

Chibani, Belattar, and Bourouis [CBB13] observed that simulation frameworks and the codebases built upon them can exhibit cross-cutting concerns such as event handling, resource sharing, and the restoring the state of a simulation run. They investigated the introduction of aspect orientation to Discrete Event Simulation (DES) frameworks to address tangling and scattering that may arise from the cross-cutting concerns they identified. They contribute a discussion of aspect-oriented programming's potential application to DES codebases, and detail the avenues available for research in the field. Chibani, Belattar, and Bourouis [CBB13] identify Japrosim [AB08; BB14] — a DES framework previously developed by the research team, which was designed to model domains through process interaction in Java — as an example of an existing framework which exhibits the cross-cutting concerns they describe.

Later, Chibani, Belattar, and Bourouis [CBB19] identified opportunities for the use of aspect orientation in simulation tooling, aiming to increase “modularity, understandability, maintainability, reusability, and testability” by applying the paradigm [CBB19]. They present a case study of an application of aspect orientation to simulation tooling by identifying cross-cutting concerns in Japrosim, a discrete event simulation framework, and propose an aspect-oriented redesign of the tool using AspectJ. Chibani, Belattar, and Bourouis describe Japrosim's existing object-oriented design, followed by aspect oriented variations of some design elements, including concurrent process management and in Japrosim's graphical animation features. A similar survey of areas in which Japrosim's source might benefit from the application of aspect orientation is presented by Chibani, Belattar, and Bourouis in an earlier work [CBB14]. In both cases, the main contribution noted is the design itself. Counting the main improvements between the presented aspect-oriented design and the existing object-oriented one is left to future work in the authors' later publication [CBB19], although a concrete implementation is linked to and some quantitative evaluation of that implementation presented in their earlier publication [CBB14]. The quantitative evaluation provides measurements based on Martin [Mar94]'s object-oriented design metrics and demonstrates a greater independence of packages in their aspect oriented version of Japrosim than in the original. However, the intended aim of aspect orientation is

Evaluating?

new sentence here
- Chibani, Belattar, and Bourouis [CBB19]'s later work presented an implementation of their

not to decouple existing packages, but to isolate those packages' cross-cutting concerns into new ones. It is therefore unclear that their quantitative evaluation achieves its improvements as a result of aspect orientation. No further discussion of their results is provided, and it is possible that the improvement is due to the rewriting necessary in their maintenance of the Japrosim source, rather than due to their use of aspect orientation specifically.

Similarly to Chibani, Belattar, and Bourouis [CBB19; CBB13; CBB14], Aksu, Belet, and Zdemir observe that there are advantages in adopting aspect orientation when developing a simulation framework [ABZ08]. Examining the DES framework Simkit, they motivate two different applications of aspect-orientation: a refactoring of the framework itself to better manage cross-cutting concerns within its codebase; and aspect-oriented tooling for use by modellers who represent cross-cutting concerns within their models. Opportunities for improvements in production and development are discussed, and some implementation notes are detailed, although no concrete implementation or evaluation is provided; the work instead proposes design alterations, and the authors “leave it as a future work [*sic*] to explore the usability and efficiency” of aspect orientation used idiomatically alongside Java’s existing reflection offerings. The existence of multiple attempts to refactor differing simulation packages with aspect orientation indicates potential for modellers in the use of aspect-oriented patterns, but the real-world utility of the techniques are omitted. As is common in aspect-orientation literature, Chibani, Belattar, and Bourouis and Aksu, Belet, and Zdemir both defer to the general claims that aspect-orientation improves modularity of cross-cutting concerns and can eliminate code smells such as tangling and scattering.³

grab a citation for Simkit

Aspect-Orientated Implementations of Simulations & Models

Research projects applying aspect-orientation to the implementation of simulation frameworks [chibani] fails to provide a case study which evaluates their technique with real-world examples. However, case studies do exist which evaluate aspect-orientation’s suitability in maintaining model source code.

Ionescu et al. identify an increased demand for computational power in simulation execution on supercomputers [Ion+09]. Existing known-good models might be unsuitable for the extreme requirements of code efficiency modellers contend with, but running the code in different environments

³Chibani, Belattar, and Bourouis do present some quantitative evaluation, but this is flawed as previously described.

requires modifications to make the code suitable in the environment. These modifications are subject to regulations and introduce risk of a reduction in quality during maintenance. The authors propose an aspect-oriented solution to the problem, where aspects modify the simulation codebase with minimal overhead. An implementation of a real-world model for disaster prevention is presented, and assessed both by comparison against an equivalent non-aspect-oriented codebase and by assessment of the aspect-oriented variant's scalability and reliability in both cluster and multi-cluster environments. They find that a comparative analysis of generated code and of their simulations in various configurations both indicate that their simulation's aspect-oriented implementation is suitable for use in disaster prevention, implying that aspect orientation could be suitable in scenarios with comparable requirements.

This work provides evidence that aspect-oriented programming is suitable for supporting modifications to simulations and models. This contrasts the lack of evidence for the paradigm's success in improvements to modularity and maintainability [Prz10; CSS04], and supports Steimann06paradoxicaempty citation's suggestion that aspect-orientation might be suitable for other uses — as well as the suggestion made by Gulyás and Kozsik [GK99] that the paradigm is well suited to the requirements of simulation & modelling codebases.

As it is a rare example of aspect-oriented case studies, the methodology employed by Ionescu et al. is important to highlight. Their evaluation measures code quality improvements, which are commonly claimed in aspect orientation research . Their code analysis makes use of significant lines of code as a core metric, which doesn't reliably reflect code quality. As Rosenberg [Ros97] explains:

“(...) the best use of SLOC is not as a predictor of quality itself (for such a prediction would simply reduce to a claim about size, not quality), but rather as a covariate adjusting for size in using another metric.”

add citations
where code
quality im-
provements
are claimed
in AOP re-
search

Although Ionescu et al. [Ion+09] evaluate code quality, the methodology employed to measure improvements is unreliable.

Improvements in code quality are those which have come under scrutiny by the critical papers reviewed in section 2.0.2. The results presented by Ionescu et al. do not satisfy critics' requests for empirical evidence of improved code quality. This does not impact their aspect-oriented models'

viability: their study demonstrates that their models can be augmented to support new supercomputing environments without lack of performance. The models described in this work satisfy that aim: their models are also evaluated in their performance. Model performance is a priority in supercomputing contexts, where execution time is financially expensive and energy-intensive. Quantitative evaluation of their models' execution time shows less than 5% slowdown compared to a non-aspect-oriented implementation. Ionescu et al. deem this a reasonable trade-off for the engineering improvements they observe.

Ionescu et al.'s application of aspect-orientation to supercomputing & disaster prevention simulations meet their performance requirements and demonstrate a modelling technique which adapts existing models for use in new environments without directly modifying pre-existing source code. This result is notable with regards the contributions presented in this thesis, which similarly aim to augment a pre-existing model without directly modifying its source code — though this thesis' targets model reuse and design simplification rather than for avoiding the regulatory overhead and financial cost of maintaining models which run on supercomputers.

2.2.3 Aspect Orientation & Business Process Modelling

Several projects within the business process modelling research community make use of aspect orientation to design modelling languages which produce less monolithic business process models [Cap+09; Sil+20] and simplify the composition of models [CM07]. Business process modelling research is relevant to this thesis' contributions, as business processes are inherently socio-technical and later chapters present tooling for and results in the modelling and simulation of socio-technical systems using aspect-oriented techniques. In addition, some research conducted prior to this thesis developed software engineering processes that are conceptually similar to business process modelling (see chapter 3). There also exists interest in modelling behavioural variance within the business process modelling community (see section 2.3), which is relevant to this thesis' concern with the aspect-oriented representation of changes to processes and modelled behaviours.

Charfi and Mezini see opportunities in integrating BPEL, an executable business-process modelling language, with aspect-oriented concepts [CM07]. This is because when BPEL systems are composed

past tense
again - the
paper has
been writ-
ten/pub-
lished - fore-
saw, ob-
served, de-

together the static nature of the logic being composed is not always appropriate for BPEL's use cases. The specific use-case examined is web service definitions, where changes affecting composition of multiple component parts can affect many areas of a final result, making modification error-prone. They specifically seek to support dynamic workflow definitions — “adaptive workflows” — which BPEL's existing extension mechanisms do not sufficiently support, but the aspect-oriented literature discusses at length (an overview of which is presented in section 2.1). Therefore, they look to construct an aspect-oriented BPEL extension. Using the case study of modelling a travel agency's web services, they create an aspect-oriented extension by first defining how such an extension would be represented graphically in BPEL's workflow diagrams. Further detail is added to arrive at a technical definition with XML representations, weaving mechanics, and eventually the construction of a BPEL dialect, AO4BPEL. The authors find that their pointcut system (which describes join points on both processes and BPEL messages), support for adaptive workflows, and aspect-oriented approach to workflow process modelling make AO4BPEL unique at the time of publication, though related AOP implementations exist in each individual area of their contributions. The work is weakened by brittle semantics around pointcuts, join points, and the temporal nature of workflow modelling. For example, they note that defining contingent behaviour — only applying an aspect conditionally, based on a trace through a simulation of a modelled system — would allow the application of advice only when model state deems this appropriate.⁴ They also call for more generally theoretical AOP research, which mirrors the requests some critics of aspect orientation research make (as noted in section 2.0.2).

what do you mean by static nature? not clear

Can you give an example

In a PhD thesis describing AO4BPEL in detail [CM06] Charfi and Mezini presented a generalisation of the notation developed for AO4BPEL, which applies to any graphical workflow modelling language. Accompanying this are some examples of its use building a framework for enforcing certain requirements of BPEL models, and use of that framework to develop aspect-oriented frameworks for enforcing security and reliability within AO4BPEL models.

this seems relevant argument so avoid footnotes?

Don't need to state pub type, just name the author

In later work, Charfi, Müller, and Mezini define a similar aspect-oriented dialect of BPMN they name AO4BPMN [CMM10], after asserting that the concerns addressed by AO4BPEL [CM06; CM07] in the field of executable process languages also apply to business-process modelling languages, and can be solved similarly. The generalised notation of aspectual workflow models presented in Charfi

⁴The contingent application of model adaptation is a motivating case for some work presented in this thesis; see chapter 3 for a discussion.

and Mezini’s thesis [CM06] are applied to BPMN to produce an aspect-oriented language specifically for process modelling, as opposed to executable business process modelling.

can you
say some-
thing explicit
about the dif-
ference and
benefits of
this?

Cappelli et al. also note that cross-cutting concerns exist in business process models, and are specifically motivated by monolithic design approaches common in business process modelling languages. Like Kiczales et al., they claim that a lack of modularity in business process models leads to cross-cutting concerns scattered throughout a model [Cap+09]. To alleviate the issue, they propose a meta-language, AOPML, which incorporates aspect orientation in a metamodel of business process modelling languages, and instantiate it within their own dialect of BPMN. Using a model of a steering committee as a case study, and separating cross-cutting concerns such as logging, the paper proposes reducing complexity and repetition graphically, thereby in a manner more in keeping with the language design philosophies of popular business process modelling languages, the design and use of which are typically graphical [Obj11; DOR95; Obj15]. They note that this is in contrast to other applications of aspect orientation in business process modelling — specifically AO4BPMN — where aspect definitions are written in XML concern not only the advice to be applied but also their relevant join points, as in general programming aspect orientation implementations such as AspectJ. In this way, the AOPML

Can you say
something
more precise
than spirit
here, e.g. ob-
jectives, in-
tentions at
least?

exhibits the spirit of business process modelling more stringently than does Charfi and Mezini’s notation for aspect-oriented workflow modelling.

The difference between Charfi, Müller, and Mezini’s approach in designing AO4BPMN [CMM10] and Cappelli et al.’s approach in designing AOPML [Cap+09] highlights design decisions taken when introducing aspect orientation in a new domain. There is an opportunity for a domain-specific aspect orientation framework to align its design with the traditions and idioms already present in models within that domain, but doing so may break the traditions and idioms which already exist in aspect-oriented approaches in other domains. Comparing the approaches of Charfi, Müller, and Mezini and

demonstrates

Cappelli et al. does surface that there may be no clear “best” design approach when blending pre-existing modelling paradigms, such as the graphical modelling languages used in business-process modelling and the abstract concepts of aspect orientation. The discussion around whether it is more desirable to adapt existing design elements of aspect-oriented frameworks to a given domain or adapt

conven-
tions, syntax
- again be
more precise

that domain’s existing modelling traditions and idioms to incorporate aspect orientation as it is used elsewhere is outside the scope of this thesis.

New concepts within the design of aspect orientation frameworks are addressed in the business process modelling community. Jalali, Wohed, and Ouyang note that aspect oriented modelling frameworks often do not explicitly model the precedence of aspect application [JWO12]. They address this limitation by defining a mechanism to be used in capturing multiple concerns as aspects, where the invocation of advice must follow a certain precedence. The aim of the work is not to propose tooling around the precedence of aspect application so much as to contribute to aspect oriented design theory, providing a notation for precedence which is broadly applicable. The precedence model is, put simply, that a mapping exists for each application of advice to join point such that the mapping defines an ordering on advice for that join point. The definition defines “AOBPMN”, a formalised dialect of BPMN supporting aspect orientation with precedence. A case study is provided where AOBPMN is instantiated within a coloured Petri net. Their study expands on existing work by research teams led by Capelli [Cap+09; Sil+20] and by Charfi [CM07], in that it develops a mature formalism for and model of aspect orientation as applied to business process modelling. However, Jalali2012aspect note that their case study is limited in scale. No tooling or evaluation of the practical benefit of their approach is provided.

2.3 Process Variance in Simulation & Data Generation

Brief intro of the section here **Brief intro of the process variance section here**

Typically, a simulation concerns a single process. This means that all expected behaviour must be included within that process; complex or contingent behaviour must be represented within it. The techniques reviewed here offer separately including some modification of a process (or represent the modifications of varied process within the simulated output [SA13; SA14]). The benefit of this approach is that possible changes to a process can be described once and applied to that process where appropriate. Process changes might describe attempts to circumvent security protocols, laziness or confusion in a human actor within the model, or random “noise” so as to produce synthetic log traces containing aberrations which mimic those found in noisy empirical datasets. In all cases, behavioural variations can be described as some alteration to a process and applied to either a model or the product of that model (datasets or log traces) to represent the same alteration introduced at an arbitrary point

add something early on giving an example of how it might be a problem?

Agree, builds well from preceding sections though

Not sure this is true - e.g. MAS?

of the simulation.

This decouples the expected behaviour in the original model from simulated behaviour, which is obtained by composing the model and behavioural variation using a given technique's method for doing so. This approach to modelling behavioural variation allows the same altered behaviour, which would otherwise be described in many disparate points in a model, to instead be written once and introduced wherever required. The observation that the same variation might appear in many areas of a model, and that the variation can be separated from the model and introduced where necessary, frames the modelling of these variations in the same way as aspect orientation frames cross-cutting concerns. The work presented in this thesis explicitly applies changes to processes and simulated behaviour as aspects in the same manner. Therefore, although this aspectual connection is not made explicit in much of the literature to date, it is important to review literature on simulation and modelling which modularises these variations; this section reviews that literature. The work reviewed is highly relevant to the contributions in this thesis, in particular because the core motivations of this field are shared by this thesis; the section therefore leads with a subsection discussing those motivations, and their relationship to aspect orientation, in detail.

2.3.1 Discussion of Variation & Motivations for Variations in Process Models

Mitsyuk et al. are motivated by the field of process mining's requirement for datasets of process logs made from well-understood process models, defined in a high-level manner [Mit+17]. They demonstrate a technique for generating event logs from BPMN models by introducing algorithms for the direct simulation of BPMN models and the collection of traces from those simulations. While their approach does not support the simulation of all BPMN concepts, notably message passing, they provide a tool which produces log traces for a BPMN model through PROM, a standard tool within the process mining community [Van+05]. This results in their technique providing high-level model simulation through already-standard tooling, meaning adopters of the technique need not rely on dedicated tooling which may not be compatible with other researchers' process mining techniques.

The algorithms presented by Mitsyuk et al. simulate processes described by BPMN models, but

remove plausibly - it can or it can't

don't include any provisions for representing variance. However, the technique could plausibly

be combined with aspect orientation techniques for BPMN as discussed in ?? [CMM10; Cap+09] to represent alternate behaviour applied contingently. Demonstrating the viability of this approach is an avenue of research beyond the scope of this thesis. However, the motivation of the work mirrors that of other research projects reviewed in this section: a need for synthetic datasets of traces through a process, for use in scenarios where empirical datasets are difficult to obtain.

Difficulties arise when obtaining real-world datasets for many reasons. For example, large empirical datasets are typically produced by organisations which would prefer some level of secrecy around their operations, making publishing those operations for the investigation of research teams unlikely. Researchers collecting these datasets describe a “lengthy process” [Don20] and explain that traces of real-world processes are hard to obtain because “higher management [can be] worried about the risks” of publishing such datasets [Don20]. Another factor contributing to the difficulty of collecting empirical datasets is that they often cannot be collected, either because there is a need to study the process before implementing it (making synthetic datasets the only option available to researchers **Find citation for empirical datasets being the only ones available to researchers, maybe for disaster prevention or similar?**), because the process is not yet fully understood (making simulation of many variants of that process useful in aiding understanding **find a citation for simulating different systems for finding an optimal one. Arguably Genetic programming & hill climbing do similar things?**), or because the dataset itself is of use to researchers, not the real-world system that produced it (such as in the case of evaluating process mining techniques [VW04; AGL98]). **Find some nice way to round this subsection out, or refactor out subsections Important to acknowledge somewhere in this subsection that synthetic data generation is a well-researched field, but that generating logs from simulations with variance is the specific area relevant to this thesis**

Has this motivation already been explained? If not re-order with below first.

2.3.2 Representing Variations in Process Models and their Outputs

Research undertaken by Stocker and Accorsi [SA13] aims to synthesise process logs which are representative of attackers’ efforts to compromise the security of a modelled system. Their research project, named “SecSY”, is an attempt to address issues arising from the difficulty of retrieving representative log traces for security-critical systems in which attacker activity is present. Logs are developed by process simulation through “well-structured” models, a mathematical property on which transfor-

refactor by re-ordering - explain the motivation first and then intro the approach of charfi

mations were previously defined by Vanhatalo, Völzer, and Koehler [VVK09]. The authors develop a tool for the simulation of a process using well-structured process models, and apply transformations to both the model before execution and the log it produces through the trace of a simulation. They

with what particular characteristics? e.g. that plausibly mimic real world logs of real security violations?

is this a however, i.e. a limitation?

further

techniques

say limited instead of relatively rudimentary?

why only presumably?

new para-graph?

of?

conclude that their tool is performant, and verify it can produce logs representing security violations by way of analysis through PROM, a popular framework for process mining, and pre-defined security constraints on their models. They note that log traces cannot be interleaved (due to a lack of parallel simulation of processes), may be incomplete (missing violations), and that mutated models and traces are not guaranteed to be sound by construction. However, they see their proposal as a necessary step in realistic data generation for business processes. A weakness of the work is that model and trace modifications are relatively rudimentary : processes can be added or removed, but complex graph transformations are presumably only permissible when representable through the composition of the mutation primitives they provide, on which there are only three for processes: swapping AND and XOR definitions of process gateways, and swapping process order. Mutations cannot be applied contingent on the state of a simulation run, for example, representing a decision taken by an attacker based on what had already happened. In later work, Stocker and Accorsi detail the technical aspects of SecSY, their tool for implementing the generation of synthetic logs which use their technique [SA13] to represent security violations security-critical business processes. A Java implementation of SecSY is described, which simulates well-structured models and applies mathematically-defined transformations on the model being simulated (before simulation occurs) and the logs obtained through simulation traces. An improvement on earlier work is that custom transformers can be written. However, a limitation of the original work remains, which is that users cannot easily dictate the degree to which variations are applied.

Pourmasoumi et al. also address the need for access to variations on business processes, though for the development of a research field, “cross-organisational process mining” [Pou+15]. Process mining can require many process logs, as does the benchmarking and evaluation of process mining techniques. Traces from business processes which are similar but not identical can produce log traces which reflect that similarity, but also reflect the variations in different instances of those processes. These log traces exhibiting variation can be used in the training and analysis of process mining tooling and techniques, which must contend with natural variation present in the execution of real-world traces. To support

the field, log trace generation from a variety of process models is therefore required. Such logs are not in adequate supply, as explained in section 2.3.1. The authors' approach to the problem is to present an algorithm for the mutation of business processes, such that simulation against variations of the business process can produce process logs reflecting those variations. Their algorithm makes use of structure tree representations of processes, which models processes as trees and permits conversion to BPMN models and Petri nets [Bui14]. Pourmasoumi et al. make use of this constraint to demonstrate that their models are block-structured, a mathematical constraint on model structure which 95% of models are shown to comply with [Li10]. Their contribution is a set of transformations on structure trees and block-structured models, and an algorithm applying these transformations to process models, and a tool which implements it built on PLG, a process log generation tool. They conclude that tools such theirs can be used to generate log traces representing process variation, in such a manner as to satisfy the requirements of the process mining research community.

Pourmasoumi et al. describe a list of transformations they explain is "not intended to be comprehensive" [Pou+15], which makes its full potential unclear. Additionally, processes their tool applies to must be block-structured. The importance of this requirement is that it limits their technique to business process models. It is not demonstrated that models of processes in other domains satisfy the condition, such as the flow of data [WC10] or behaviour of human or technical actors in socio-technical systems [WS18a]. Finally, the tool is limited by its lack of capacity to represent variations which are applied contingent on system state. A use case for modelling behavioural variance is to model changes which are impossible to anticipate from the vantage of a modeller, such as a socio-technical system's human actors' mistakes, security exceptions and violations, or corrective actions to mitigate undesirable system state. Pourmasoumi et al.'s tool produces variations on a process model, but modelling behaviour which is expected to vary within iterations of the *same* model is outwith the scope of their project. **Shugurov & Mitsyuk's work should follow this. REVIEW IT.**

be explicit
- on what.
Also start
with the
overall thrust
of the work,
not the specific contribution

as per first
comment
on this paragraph - what technique?

Machado et al. note that there are operational costs to the inefficient modelling of business processes. Specifically, they note that processes can be replicated across automated processes , and failure to identify such scenarios give rise to these operational costs. This work's core motivation is that the representation of variation in process models would allow for the capture of a replicated process once, with instances of similar processes described as deviations from that captured blueprint. On this basis,

unclear here

Machado et al. extend BPMN to support the notion of individual processes as transformations of an underlying process, i.e. that a given process can sometimes be expressed as a deviation from some pattern, and is therefore define-able as the composition of that pattern and a variation upon it. Their approach is illustrated through two small, broadly similar business processes initially modelled in BPMN and then represented in Haskell, allowing the authors to demonstrate their representation of variability as process deviation with realistic examples. While the work presented makes no empirical evaluation of their technique, Machado et al. note that their industrial partner responded positively to the research presented in this publication and that further technical improvements are to be made (support for around advice, and for quantification). They also express an intent to conduct a real-world evaluation in the HR domain. While we are unaware of any real-world evaluation undertaken by this research time to date, some formal proofs that the transformations their tool supports are always

have been
developed

well-formed [MAG12].

2.4 Discussion

This sentence was salvaged from the simulation & modelling section. It's OK but was waffly there. Maybe it can fit in this discussion somewhere?

This thesis draws on the idea that, in response to Steimann asking what aspect orientation is good for, Gulyás and Kozsik would seem to answer, “simulation & modelling”.

Aspect orientated programming is designed to permit highly modular software engineering in scenarios where cross-cutting concerns are identified, by isolating them as separate modules [Kic+97]. The aim in employing aspect oriented programming is to reduce tangling, where a cross-cutting concern is intertwined within a program's main concern, and scattering, where the same cross-cutting concern is re-written at many points in a program's source. Aspects which modularise that concern can be written once, separately from a program's main concern, and re-introduced to each point in a program to which the aspect relates by way of weaving. An aspect orientation framework must therefore be able to quantify the points at which the aspect applies [FFN00]. Aspects specify both advice (the implementation of its associated cross-cutting concern) and the join point defining where in a target program that advice should be woven. Aspect orientation therefore implies that the source aspects are applied to are oblivious to their application [FFN00].

In theory, the design of the paradigm is such that it should be expected to increase modularity in the

Worth observing that Ionescu et al. [Ion+09] actually produce a case study of using AOP and their code quality evaluation uses

software applying it [Kic+97; FFN00] **more citations** , and its proponents often claim this modularity as a benefit of the aspect-oriented approaches of their research [GS04; CM07; Cap+09; JWO12; CBB13]. However, its critics question the reasoning around these benefits, and note that there is little empirical study into whether aspect oriented programs truly benefit as a result of this modularity [CSS04; Ste06; Prz10].

One appropriate application area may be in the representation of behavioural variation in simulation and modelling. The application of aspects to models is already well-studied [ABZ08; CBB13], particularly within aspect-oriented business process modelling [CM07; Cap+09; JWO12], where modelling behavioural variation has also seen some prior research [Mac+11; SA13; Pou+15; Mit+17]. Outside of business process modelling, aspect orientation would reasonably be expected to support the development and observation of models themselves [GK99].

Research opportunities at the intersection of aspect orientation and the modelling behavioural variation occur because behavioural variation is an example of a cross-cutting concern. Changes to expected behaviour such as laziness, boredom, confusion or misunderstanding can impact many parts of a process in a socio-technical system, but modelling the variation caused by any one of these requires similar alterations to behaviour in many disparate parts of the model they occur within. Behavioural variations are therefore both scattered and tangled, and constitute a cross-cutting concern which might be well suited to modularisation in aspects.

Existing aspect orientation techniques and behavioural variation modelling techniques are ill-equipped to take advantage of this alignment. That behaviour changes when it varies is tautological; however, changes supported by existing aspect orientation techniques weave advice before, after, or around their join points, and therefore do not alter the definition itself. In some systems, some variations may be representable as additions inserted before and/or after some other behaviours, but such techniques are unsuitable for representing modifications of the target behaviour itself, or behaviours which should be omitted instead of added. Additionally, join points available to an aspectual programmer may not be granular enough to permit representing the changes they require in such as system, and aspect orientation's principle of obliviousness opposes the modification of target code to make new join points available. Techniques which would directly rewrite target source are typically extremely low-level, and therefore ill-suited to most modelling applications [KH98].

Other techniques which permit defining changes to processes at a high level may allow a modeller to *describe* intended changes (such as the high-level annotations supported in AOPML [Cap+09]), but such techniques are intended for human interpretation, not machine execution for simulation purposes. These techniques permit describing behavioural variation within another process, but only by virtue of the flexibility of natural-language annotation, and are therefore unsuitable for simulation and modelling purposes.

Such techniques also lack executable notion of “state”. Real-world behavioural variance can often be contingent on the environment an actor exists within. While variations such as security breaches might be predictable (by identifying weaknesses in existing processes, for example), variance in socio-technical systems often occurs in the behaviour of human actors. This might be in response to a degraded mode [JS07], where behaviour naturally drifts to a functional — but undesirable — state, or due to an individual making a mistake, forgetting procedure, or being in a state which alters their behaviour, such as tiredness or drunkenness [OLe20]. A framework for modelling behavioural variation using aspects should therefore apply that aspect to a simulated system contingent on the state of that system at a given point in time. High-level modelling technologies such as BPMN and OPM are executable [Mit+17; DOR95], but it is not trivially evident that executable versions of these technologies are compatible with aspect-oriented modifications of their modelling language [CMM10; Cap+09]. As noted, low-level program transformation technologies are also unsuitable. Techniques for applying variations to models exist, but are unsuitable for simulation (and therefore cannot represent application based on temporal state) [SA13], produce individual models representing each possible instance of a variation [Pou+15], do not support the dynamic weaving of aspects for contingent behaviour [Mac+11], or attempt to represent the changes one would expect in the output of a simulation by executing an unmodified simulation and amending its output directly [SM14]. None of these techniques are suitable for representing a behavioural change which is applied contingent on state. Incidentally, these techniques for modelling behavioural variation also lack support for the alteration of a process definition, or changes “inside” a process definition, as discussed earlier, which also makes them unsuitable for simulated behavioural variation.

This is good, but don't be speculative like this. Draw a conclusion

Should a tool exist which dynamically weaves definitions of behavioural variations for contingent application and which is capable of expressing changes within a join point rather than before or after

it, a challenge would arise as to demonstrating the benefits the tool achieves. Contingent application of behavioural variation, and the ability to define changes to processes specifically, would fulfil the opportunity in marrying aspect orientation and modelling behavioural variation, but the benefit offered by such a tool is unclear. The introduction of oblivious modification to a model may break its representation of its real-world analogue, making such a model difficult to reason about. Added to this is the complexity of the tool's capacity to rewrite any join point's definition. Though aspect oriented literature often lacks case studies demonstrating the benefits of the approach, it is particularly important to investigate whether such a tool could produce realistic models, and whether the expected benefits of aspect orientation as applied to the model hold in practice. In particular, would engineers benefit from the modularity afforded by isolating behavioural variation into an aspect, and would the resulting aspectual modules be re-usable when modelling other systems?

2.4.1 Research Questions

Not sure exactly how to write this. Eeep!

Chapter 3

Prior Work

An implementation of the main tool developed and used in this thesis, “PyDySoFu”, predates this thesis. As context for the contributions in this thesis, this chapter will describe the state of the project before the presented research was undertaken. Motivations for the tool’s original development are described in section 3.1, which are followed by its design and implementation in section 3.2, and that of related tooling for experiments and simulations in section 3.3. The chapter concludes with a description of the research undertaken using these tools in section 3.4, as some results in the representation of behavioural variance using aspect orientation were found using these tools which predate this thesis and offer important background for the research undertaken in it.

3.1 Motivations in originally implementing PyDySoFu

The original incarnation of PyDySoFu was developed with different motivations than those outlined in chapter 2. It is therefore important to elucidate the context in which it was designed and developed.

PyDySoFu was originally developed for the representation of behavioural variance in sociotechnical systems, and was first produced as a proof-of-concept; the core contribution was one of tooling. It was developed for use in Python codebases, by virtue of the language’s widespread use and its flexibility in its modelling of data and process.

The original version of the tool was to be applied to models of behaviour in socio-technical systems, where individual actions were represented as functions. Actions which could be decomposed further into more granular actions were to be defined as functions with calls to invoke their more granular counterpart functions. Invocations of low-level behaviours would implement some change to an environment in the model which its modelled behaviour would be expected to incur. Invocations of high-level behaviours, containing the invocations of lower-level behaviours they compose in the model, would therefore apply the combined effect of the collected behaviours they represent. A benefit of this approach to modelling behaviour was that high-level behaviours could implement the “flow” of a behaviour. For example, a behaviour which would be modelled in a flowchart as having some loop could be modelled analogously in the method described through use of primitive control flow operators in Python, such as `for` and `while` loops.

Another benefit of this approach is that the behaviours modelled have a predictable structure which is amenable to metaprogramming. A low-level behaviour’s affect could be changed by changing the function definition; more structural changes could be made by altering the flow of less granular behaviours. A simple high-level behaviour containing a series of function invocations (modelling an ordered list of steps in the socio-technical system) can be represented as a literal list of function calls. The contents of such a list is trivially modifiable. Removing an item from a list or truncating it at a certain length, for example, are both achievable in a trivial manner using high-level languages such as Python. Notably, many behaviours can be conceived of which could be represented as high-level behaviours but would not be amenable to a simple list of more granular behaviours, such as a behaviour with a looping quality.

With a mechanism to rewrite either an implementation of a behaviour or a collection of behaviours (in the less granular functions mentioned), modelling in such a fashion could therefore lend itself to semantically simple metaprogramming that could represent real-world variations in behaviour. However, largely for reasons discussed in ??, metaprogramming for representing realistic behavioural variations in socio-technical simulations should be able to take advantage of system state. Many real-world behaviours are contingent on environmental state. Real-world actors in socio-technical systems might become tired after lots of work, or proportionally to time of day within a simulation. Therefore, the metaprogramming as described should be performed during runtime, for which no suitable candidate

was available. PyDySoFu was developed to fulfil this requirement, so that behavioural variance in socio-technical simulation could be modelled as described and subsequently studied.

3.2 Original PyDySoFu Implementation

To disambiguate the improvements made to the original PyDySoFu implementation in the tooling contributions of this thesis — and to explain the fundamental concepts involved in both implementations’ approaches to aspect orientation and the application of behavioural variance — the implementation of the original PyDySoFu tool, which predates the work presented in this thesis, is described here.

3.2.1 Weaving mechanism

The original PyDySoFu implementation made use of a weaving mechanism which could be categorised as “total weaving” in the parlance of Chitchyan and Sommerville [CS04]: hooks to apply advice are woven into every possible join point. The library was implemented in Python, which offers a flexible object model PyDySoFu is able to take advantage of in order to weave its hooks.¹

Python’s object model has three key properties which the original implementation of PyDySoFu takes advantage of:

- ① Everything in Python is an object, including types, functions, and classes
- ② Objects are, in essence, implemented as a dictionary² with string keys. All attributes of an object — such as a method on an instance of a class — are values in this dictionary, and their identifiers are the string keys of the dictionary. In essence, `someObject.val` is notionally equivalent to `someObject.__dict__['val']`, though the subtleties of this mechanism will be explained later.
- ③ Operations on objects are handled by “magic methods”, which are specifically named methods

¹The weaving mechanism of PyDySoFu was eventually factored out into another library, *Asp* [WS23]; as PyDySoFu was mostly used before this refactoring, particularly in case studies (see section 3.4), the weaving mechanism will here be described within the context of PyDySoFu specifically.

²Python’s name for what other languages might call a *map* or *hashmap*.

on objects which Python calls for fundamental behaviour in the language. For example, `objA == objB` is interpreted by Python as `objA.__eq__(objB)`. Other built-in behaviours in Python have similar reserved method names which represent the implementation of some behaviour. These methods can be overridden, or specified by a programmer if they wish.

PyDySoFu weaves aspect hooks into classes by taking advantage of these three properties of Python. At a high level, PyDySoFu operates by replacing the `__getattr__()` method of a class object with a custom one. `__getattr__()` is the magic method responsible for performing lookups in an object's underlying dictionary. The replacement `__getattr__()` also looks up attributes in the relevant object's underlying dictionary, but it also searches for advice to be applied when performing these lookups, and applies any advice it finds. The replacement `__getattr__()` method represents the aspect hooks woven by PyDySoFu.

Hooks are applied to a class by way of an invocation to a function, `fuzz_clazz`, which takes a class as a parameter and weaves aspect hooks into that class [WS18c; WS23]. `fuzz_clazz` replaces the `__getattr__()` method of the class with a new function object which it constructs. The replacement function object injects aspect hooks to callable attributes, thereby discovering aspects, applying them, and invoking an attribute which is the target of advice within a wrapper of the attribute itself.

The replacement `__getattr__()` method makes a call to the class' original `__getattr__()` method to retrieve an attribute when required. If this attribute is not a function or method, it is returned by the woven `__getattr__()` function and the program affected class behaves as if it was never altered. However: if an attribute to be retrieved is a method or function, a new function is constructed and returned. This function looks up and applies advice to the attribute originally retrieved by the program, and contains a reference to the original aspect to enable its execution. This wrapping function is therefore the core of the weaving process: a replaced `__getattr__()` method injects the aspect hooks which constitute the aspect-orientation framework. Advice to be run before, around, and after a join point were implemented as invocations of advice functions before a target was executed for "before" advice, after a target was executed for "after" advice, and with an around function being run with a reference to the advice's to invoke at its own discretion.

3.2.2 Applying Process Mutations

The development of PyDySoFu was intended to support simulation & modelling research by providing a way of applying program modifications at runtime. It sought to fulfil the needs of a program which determined that a (potentially non-deterministic) change to the behaviour it modelled was required, and enable the program to apply that change mid-process in an aspect-oriented fashion. Aspect orientation libraries typically support advice woven before, around, or after a join point; modifying the join point itself essentially allows changes inside its definition, introducing a fourth type of weaving. PyDySoFu achieves this through a special type of “before”-style aspect, which it calls a “fuzzer”.

Fuzzers implement transformations on abstract syntax trees. They are implemented as functions which receive a list of AST objects representing the definition of a function which satisfies a fuzzing advice’s join point, and return another list of AST objects, which replace the target function’s definition. Any transformation resulting in a valid AST is permitted. A code snippet demonstrating the implementation of this process is shown in fig. 3.1.

As can be seen in fig. 3.1, fuzzing aspects are implemented by way of “prelude” advice (PyDySoFu nomenclature for advice run *before* a target invocation, inspired by early work on TheatreAG, which is discussed in section 3.3.2). Prior to the target being run, the fuzzer replaces its underlying code object with one produced by compiling a fuzzing aspect’s returned AST objects. Replacing the underlying code object with a modified one allows the fuzzer to define arbitrary modifications to the definition of the function, thereby achieving runtime metaprogramming.

3.2.3 Limitations

The original implementation of PyDySoFu was intended to demonstrate the feasibility of runtime metaprogramming and its potential in simulation and modelling. However, its design presents limitations.

There is an overhead involved in running the wrapped function for every invocation of `__getattr__()`, and also in running aspect hooks for all possible join points, even when those hooks are not targets

```

1 class FuzzingAspect(IdentityAspect):
2
3     def __init__(self, fuzzing_advice):
4         self.fuzzing_advice = fuzzing_advice
5
6     def prelude(self, attribute, context, *args, **kwargs):
7         self.apply_fuzzing(attribute, context)
8
9     def apply_fuzzing(self, attribute, context):
10        # Ensure that advice key is unbound method for instance methods.
11        if inspect.ismethod(attribute):
12            reference_function = attribute.im_func
13            advice_key = getattr(attribute.im_class, attribute.func_name)
14        else:
15            reference_function = attribute
16            advice_key = reference_function
17
18        fuzzer = self.fuzzing_advice.get(advice_key, identity)
19        fuzz_function(reference_function, fuzzer, context)
20
21
22 def fuzz_function(reference_function, fuzzer=identity, context=None):
23     reference_syntax_tree = get_reference_syntax_tree(reference_function)
24
25     fuzzed_syntax_tree = copy.deepcopy(reference_syntax_tree)
26     workflow_transformer = WorkflowTransformer(fuzzer=fuzzer, context=
27         context)
28     workflow_transformer.visit(fuzzed_syntax_tree)
29
30     # Compile the newly mutated function into a module, extract the
31     # mutated function code object and replace the
32     # reference function's code object for this call.
33     compiled_module = compile(fuzzed_syntax_tree, inspect.getsourcefile(
34         reference_function), 'exec')
35
36     reference_function.func_code = compiled_module.co_consts[0]

```

Figure 3.1: A code snippet from the original PyDySoFu implementation, implementing Fuzzing aspects and applying fuzzing to a function definition.

of advice at a given moment. When an attribute is the target of advice, aspects are discovered and applied. However, aspects are discovered by lookup within the scope of the function creating a replacement `__getattribute__()` method. This design requires multiple instances of advice weaving to create multiple replacement `__getattribute__()` calls, all of which are invoked on any attribute lookup. A single target of advice which has multiple pieces of advice woven therefore incurs a performance penalty for every piece of advice applied, which must be incurred when any attribute is looked up on the target's class. If the join point defining the target may apply to many classes, each class must incur the same penalty, even if none of their attributes are targets of advice in practice. The original library was developed to demonstrate the feasibility of the idea underlying PyDySoFu (runtime metaprogramming), but the weaving mechanism implemented left room for improvement. A robust implementation with attention paid to reducing this overhead is introduced in chapter 4.

Similarly, an additional overhead is incurred by a lack of caching of the modifications fuzzers make to function definitions. One can envisage a need for runtime metaprogramming which produces different function definitions at different times: an example could be modelling different degrees of degraded modes introduced to an actor's behaviour in safety-critical systems research [JS07]. One can also envisage no such need: an example could be minor temporary modifications otherwise permanently made in program maintenance, such as to constants within a function definition, to the format of a function's return value, or adding control flow which exits a function early on a termination condition. The requirement is a product of the tool's use case in different scenarios. In scenarios where the same modification is to be made every invocation, a fuzzer need only be run once; optimisations enabling the caching of fuzzing aspects' effects would provide better performance in use cases where such a feature is appropriate.

Other aspect orientation frameworks offer support for other types of advice. Handi-wrap and AspectJ both support features related to the processing of exceptions thrown by a program [BH02a; Kic+01] and these features have inspired work into improved exception handling in object-oriented systems [MD11]. However, this version of PyDySoFu offers no direct support for exception handling. Opportunities to support the feature were therefore available for future revisions of the library to capitalise on; such a revision is also presented in chapter 4.

A final limitation is that the weaving technique this early of PyDySoFu used is incompatible

with Python3, as replacing `__getattr__()` is not possible in Python's newer version. It was determined that a tool which was of practical use to the simulation and modelling community should be produced which would remain useful to future researchers making modern models; as the existing weaving technique lacked performance, an opportunity presented itself for a complete redesign. The resulting new design is presented in chapter 4 which makes use of a new weaving technique.

3.3 Additional Simulation Machinery

Other related projects developed tooling for sociotechnical simulation & modelling. Fuzzi-Moss was a project collecting a library of standardised behavioural modifications for use in sociotechnical simulation & modelling; Theatre_AG was a project offering a model of time against which actors within sociotechnical simulations & models could act. While these projects ultimately were not used in producing the contributions of this thesis, they are outlined here as relevant to the original PyDySoFu project as they were originally developed as a suite of simulation & modelling tools to be employed together.

3.3.1 Fuzzi-Moss

Fuzzi-Moss³ was a library of standard behavioural variations written as fuzzers to be applied by PyDySoFu [SW16]. It was primarily created for use in a model of the impact of inconsistency in teams' executions of software engineering methodology, which is discussed further in section 3.4.

Fuzzi-Moss contained utilities and fuzzers such as:

- Several probability mass functions representing chance of unexpected behaviour given a length of time and an actor's...:
 - conscientiousness, a lack of which would increase chance of behavioural adaptation due to lack of effort;
 - concentration, a lack of which would increase chance of behavioural adaptation due to

³A backronym for "Fuzzing Models of sociotechnical simulations"

```

1 def missed_target(random, pmf=default_distracted_pmf(2)):
2     """
3     Creates a fuzzer that causes a workflow containing a while loop to be
4     prematurely terminated before the condition
5     in the reference function is satisfied. The binary probability
6     distribution for continuing work is a function of
7     the duration of the workflow, as measured by the supplied turn based
8     clock.
9     :param random: a random value source.
10    :param pmf: a function that accepts a duration and returns a
11    probability threshold for an
12    actor to be distracted from a target.
13    :return: the insufficient effort fuzz function.
14    """
15
16    def _insufficient_effort(steps, context):
17
18        break_insertion = \
19            'if not self.is_distracted() : break'
20
21        context.is_distracted = IsDistracted(context.actor.clock, random,
22            pmf)
23
24        fuzzer = \
25            recurse_into_nested_steps(
26                fuzzer=filter_steps(
27                    fuzz_filter=include_control_structures(target={ast.
28                        While}),
29                    fuzzer=recurse_into_nested_steps(
30                        target_structures={ast.While},
31                        fuzzer=insert_steps(0, break_insertion),
32                        min_depth=1
33                    )
34                ),
35                min_depth=1
36            )
37
38        return fuzzer(steps, context)
39
40    return _insufficient_effort

```

distraction

- A `missed_target` fuzzer, which terminated a `while` loop early if activated via a probability mass function of an actor's propensity for negligence. As an example of a Fuzzi-Moss fuzzer, this is shown in ??.

Consider moving some code snippets like those of Fuzzi-Moss, which aren't necessarily that important in the grand scope of the thesis, to an appendix.

- An `incomplete_procedure` fuzzer, which truncated the steps⁴ taken by an actor if activated via a probability mass function representing an actor's propensity for distraction

Plans were also made for fuzzers representing an actor "becoming muddled"⁵ and making mistakes in decision-making, but neither have been completed at time of writing. A discussion around the revival of this project in the context of the PyDySoFu rewrite presented in chapter 4 is given in ??.

3.3.2 Theatre_Ag

Theatre_Ag ("Theatre") is a project defining a model of time against which actors in sociotechnical models & simulations can act [Sto23]. In the project's overview, it describes itself as...:

"Theatre_Ag is a workflow oriented agent based simulation environment. Theatre_Ag is designed to enable experimenters to specify readable workflows directly as collections of related methods organised into Plain Old Python Classes that are executed by the agents in the simulation. All other simulation machinery (critically task duration and clock synchronization) is handled internally by the simulation environment.

The central metaphor underlying Theatre's model of timing is theatrical: actors in a simulation or model are members of a "cast" (a collection of actors) who enact a "workflow" (simulation steps) in a "scene" (domain model within which the actors interact). Central to the library is its clock: tasks are given durations, and a clock which synchronises all agents' position in time ticks to complete different tasks. The theatrical model theatre introduces is the context for PyDySoFu's nomenclature for its types of advice: "prelude" advice happens before a task and "encore" advice is invoked afterward, as a prelude and encore would be in a literal theatre.

⁴Represented by lines of code

⁵A behavioural variation caused by confusion.

Theatre has been used as the environment in models of TCP/IP, algorithmic trading, the spread of disease [OLe20], and the impact of behavioural variation in software engineering methodologies as described in section 3.4.

3.4 Example Studies using PyDySoFu for Behavioural Simulation

The viability of encoding behavioural variations as aspects using PyDySoFu has been demonstrated in earlier studies [WS18a; OLe20]. The study modelled software engineers working to different methodologies of software engineering: waterfall, in which requirements are gathered, software is developed to meet requirements, quality assurance steps are undertaken, and the resulting software is delivered to customers; and TDD, where the development of tests for quality assurance precedes the development of features. The study sought to investigate whether, when software engineers were working suboptimally, there was a difference in the rate of bugs introduced to a program developed under each methodology.

Maybe I should include something on Aran's masters dissertation too, rather than just citing it?

The study began with a “naive” model of software engineers following each paradigm, developed in Python using PyDySoFu, Theatre, and Fuzzi-Moss. Engineers would produce “chunks” of code, which could contain bugs. In quality assurance, engineers were modelled as attempting to identify bugs in different areas of the codebase, fixing them if they were discovered. Developers could commit chunks of code toward features identified through requirement engineering, which were eventually completed, but could potentially contain undiscovered bugs within chunks of code.

This model was then augmented aspectually using PyDySoFu. Distraction was represented through the truncation of functions representing workflow steps. Developers were modelled with different levels of distraction, affecting a probability mass function (PMF) which would activate when a developer was modelled as being distracted in a given moment. If the PMF activated, the workflow step invoked at that moment was truncated using PyDySoFu. The model showed that developers following the TDD methodology could successfully complete a larger number of features on average than those following waterfall, concurring with the prevailing consensus on the two methodologies.

In replicating the community understanding of the model, the paper demonstrated the feasibility

Add a citation for TDD vs agile! Citations are missing from the copy of the GAISE

of aspectually augmenting modelled behaviour: the simulation took a naive model with no capacity for analysing errors, and introduced new features of the model supporting an avenue of investigation otherwise impossible with the methodologies represented by the naive model. That the resulting simulation matched the expectations existing within the community gave confidence that the tool could be used to build realistic simulations where some features of a model could be separated from its core codebase.

3.5 Discussion

The existing case study employing PyDySoFu demonstrated that realistic model features could be separated from their core codebase, and gave credence to PyDySoFu's use as a tool for aspectually augmenting models with behavioural variance. It also left many research questions unanswered and

tooling flaws unsatisfied, however:

Does this want to be a list? Consider reworking to a paragraph.

- Aspects were believed to be “realistic” as they represented the expected outcomes of the simulation. However, no real-world data was used to corroborate the claim, and it was unclear that aspectually augmented behaviour could capture the variations present in real-world human behaviour.
- These aspects also demonstrated variations which were identifiable in the emergent properties of a system (for example, mean time to failure of a software system under development, or successfully completed features). The variations applied to individual developers might have poorly modelled individual behaviour, but produced accurate emergent properties of the system individual developers acted within.
- As discussed in section 3.2.3, PyDySoFu's implementation at the time was a proof-of-concept which, while successfully demonstrating the potential of aspect-oriented runtime metaprogramming, was also inefficient, feature-incomplete, and lacked compatibility with modern software engineering tooling.
- Models of distraction were adopted from the common library provided by Fuzzi-Moss. However, this model was not applied to other codebases. It remains unclear that Fuzzi-Moss' model

of distraction is broadly applicable in other projects: different models of distraction might be required by different researchers. Further, a model of distraction which realistically represents the behaviour of an individual (rather than the emergent properties of the system that individual acts within) might not apply to other systems the individual acts within. Briefly put, The portability of aspectually modelled behavioural simulation has not been investigated, and literature within the aspect orientation community lacks evidence to support a belief in their portability [Prz10; CSS04; Ste06].

This left opportunities to improve both the tooling offered for aspect-oriented runtime metaprogramming, and the evidence supporting its use to encode behavioural variations in sociotechnical systems. Improvements to the tooling follow in chapter 4; later chapters propose — and discuss the implementation of — real-world systems which are suitable for modelling using PyDySoFu in chapter 5, and a study of those systems using aspectually augmented models in chapter 6 and chapter 7.

Chapter 4

Rewriting PyDySoFu

Check the most up-to-date pdsf implementation — is it actually on that backup drive?

The work undertaken in this thesis required in improved implementation of old tooling. When previously used, PyDySoFu was a proof of concept which could feasibly produce scientific simulations, but was implemented in a manner which was not optimised for speed (making it a burden for large simulations), lacked granularity in the application of its aspect hooks (hooks could only be applied to entire classes), and most importantly, did not work with Python3 (Python2 support officially ended during this PhD).

This chapter briefly outlines the new implementation of PyDySoFu, discusses improvements made to design and performance, and explains some contributions made to the design of aspect orientation frameworks which addresses some core issues raised with the paradigm . Consider adding references to the sections through this PDSF chapter, depending on how beefy it becomes...

Rewrite chapter outline after the structure of the chapter is known to be OK.

4.1 Requirements for Change

As time wore on with PyDySoFu's original implementation, it became increasingly clear that a rewrite was required . PyDySoFu grew out of an undergraduate project, and accrued technical debt as a result of being written under extreme time constraints with little experience. On revisiting, and

Too informal

sloppily put

on reflecting on other aspect orientation frameworks (as discussed in section 2.1 and [CS04]) and the use previously found for PyDySoFu (see [WS18a; WS18b]), it was clear that there were a series of improvements which could be made in the process of rewriting the tool:

→ Before this body of work, PyDySoFu made use of techniques for applying aspect hooks which did not translate to the changes Python 3 made to its object model. In particular, Python 3 changed its underlying object model, using a read-only wrapper class that made the replacement of `__getattr__` impossible via the previous route.

→ PyDySoFu's original implementation made no serious accommodations for efficiency. It could be seen as the "total weaving" described by Chitchyan and Sommerville in [CS04], and it was not possible to provide additional options to ensure that aspects could be as efficiently woven as possible at runtime given a particular use-case .

Could be less
sloppy

→ The original PyDySoFu implementation wove onto a *class*, meaning that even properties of the class which were not considered join points were still affected by the weaving, even if in a minor way . Because `__getattr__` retrieves all attributes including special builtin attributes and non-callable attributes, these are also returned via the modified implementation of `__getattr__`, incurring an overhead, albeit small, for all attribute resolutions instead of a desired subset.

Could be less
sloppy

→ The original PyDySoFu implementation made no accommodations for scenarios where fuzzing of source code was applied in a "static" manner. That is to say, where a deterministic modification to source is woven as advice, instead of dynamically modifying source code, the same modification would still be made every time the target attribute was executed, unless caching of results was specifically managed by the aspect applying the change. No optimisations were made pertaining to this, but compilation and abstract syntax tree editing have the potential to be PyDySoFu's most expensive operations.

→ Unlike other aspect orientation frameworks such as AspectJ [Kic+01], join points could not be specified by pattern. Instead, each individual join point must be supplied as a Python object. This means that, while the target attributes are still oblivious to the advice applied to them, the application of that advice could not be written obliviously.

As a large number of requirements were left unfulfilled by the original implementation of PyDySoFu, a new implementation satisfying them was deemed necessary.

4.2 Python3-compatible Weaving Techniques

Replacing `__getattr__` on the class of a targeted method was no longer viable in Python 3. A replacement method therefore had to be found. For clarity: replacing `__getattr__` allowed for hooks to be woven (at runtime) into likely future targets for advice. These hooks would then discover and manage the execution of advice around each target. Because advice can be run before and around a target, and dynamic weaving implies that advice could be supplied or removed at any time, we look to intercept the calling of any target, and manage advice immediately before execution. So, the task at hand is to find a method of attaching additional work to the calling of any potential target, before that target is executed. We refer to code woven around a target which manages applied advice as *aspect hooks*.

4.2.1 Abandoned techniques

Rather than “monkey-patching”¹ a new version of `__getattr__` with hooks for weaving aspects, the rewritten method could be patched to the object itself at a deeper level than used in the original PyDySoFu implementation. This would make use of Python’s `ctypes` api to patch the underlying object. Similar work has been done in the python community in a project called `ForbiddenFruit` [con21]. Efforts were made to add the required functionality to `ForbiddenFruit` — patching `__getattr__` directly on the object, or “cursing” it in `ForbiddenFruit` jargon — but this was abandoned as the underlying mechanism is particularly unsafe, Python API changes could render the work unusable in future versions easily, and the implementation would only work with particular implementations of Python (for `ctypes` to exist, the Python implementation must be written in C). Community patches existed for cursing `__getattr__` which did not work,

¹Monkey-patching is the practice of making on-the-fly changes to object behaviours / definitions by taking advantage of language properties such as flexible object structures. Common examples of these structures are objects literally being maps from string attribute / method names to the associated underlying value, as in Python and Javascript. Monkey-patching makes use of these simple structures by replacing values such as the function object mapped to by the original function’s name in the dictionary, effectively changing its behaviour. This is the method by which PyDySoFu originally replaced `__getattr__` on a class object.

and attempts proved challenging, indicating that this would also be complicated to maintain over time. There are also efficiency concerns with this technique depending on its use: weaving advice around a function would mean monkey-patching the built-in class of functions, which would incur an overhead from running aspect hooks on every invocation of every function.

Another approach involved making use of existing Python functionality for interrupting method calls. As PyDySoFu wraps method calls at execution time, what is required is to add functionality to the beginning and end of the execution of a method. Python has built-in functionality for implementing debuggers, profilers, and similar development tools, which provides exactly this functionality, as debuggers must be able to — at any point during execution marked as a breakpoint — pause a running program and inspect call stacks, the values of variables, and so on. As a result, the method `settrace()` allows a developer to specify a hook providing additional functionality to a program. Making use of this also has issues in our case. Most significantly, `settrace()` catches myriad events in the Python interpreter which PyDySoFu may not need to concern itself with, incurring significant overhead. In addition, use of the function overrides previous calls to it, meaning that any debuggers used by a user of PyDySoFu would be replaced with PyDySoFu's functionality, which was deemed untenable. However, it is worth noting that the technique could work in theory, and if future versions of Python allow for multiple trace handlers being managed by `settrace()`, this could provide an interesting approach when implementing future dynamic aspect orientation frameworks.

4.2.2 A viable technique: import hooks

A final available technique was to continue to monkey-patch hooks to discover and weave aspects, via an alternative method which did not make use of `__getattr__`. This approach would change the use of PyDySoFu slightly to make a compromise between performance and obliviousness of aspect application: when importing a module targeted for aspect weaving, methods which are potential weaving targets are invisibly monkey-patched with a wrapper method with a reference to the original² and hooks to detect and run dynamically supplied advice.

An important note for discussing the implementation of PyDySoFu is that almost all Python

²Necessary to run the originally targeted method.

functionality operates by use of its “magic methods”. “Magic methods” are methods beginning and ending with two `_` characters. The Python language documentation specifies sets of magic methods and their required function signatures which are used internally to implement functionality; for example, any object with the method `__eq__()` defined can be compared against using the `==` operator, and the `__eq__()` magic method is run to determine the outcome of the operator. Magic methods support more than operator overloading. For example, anything which defines `__len__()` and `__getitem__()` is treated as an immutable container, and adding `__setitem__()` and `__delitem__()` makes that container mutable. Any class defining `__call__()` is treated as a callable object (not unlike a function). More can be found in Python’s documentation[VD09], although more focused guides exist in the Python community [con16]. For the purposes of implementing import-based aspect hook weaving, magic methods have the affect of making the language ideal, particularly as an environment to implement dynamic aspect orientation. Python’s functionality for importing modules is managed by `builtins.__import__`, which receives module names as strings and handles package resolution. By monkey-patching the import system, modules can be modified during the process of importing. As this technique allows for control over where aspect hooks are applied, PyDySoFu can target only function and method objects to apply aspect hooks to, avoiding the overhead its previous iteration introduced when applying hooks to all attribute lookups including non-callables, such as variables or `Class` objects.

Monkey-patching `builtins.__import__` is as simple as replacing the function object with a new one, which has the effect of changing the behaviour of Python’s `import` keyword: because all Python functionality relies on magic methods implicitly, its behaviour can be altered in this way. However, our intent is not necessarily to manipulate *all* modules, but a subset of imports specified by a modeller as suitable for manipulation. If all invocations of `import` wove hooks into modules, including those made in the process of importing packages, an unnecessary overhead would be introduced when invoking any callable in any module. Therefore, it is important to have a mechanism to enable and disable the weaving of aspect hooks for a given `import` statement (effectively, to enable and disable PyDySoFu’s modified import logic).

This can be achieved through another use of magic methods in a manner which also makes clear to a modeller exactly where aspect hooks are being applied: making use of Python’s `with` keyword. A

```

1 | with AspectHooks():
2 |     import mymodule

```

Figure 4.1: Example of importing a module, `mymodule`, using PyDySoFu’s import hook design.

more technical discussion on the process of weaving, and the nature of the hooks applied, follows in section 4.3.

4.3 Import Hooks

4.3.1 Implementing import hooks

We are interested in manipulating `builtins.__import__` only when imports are made which introduce modules containing prospective join points; a developer might import many modules, which do not all require aspect hooks to be introduced. We enable this new import behaviour with a syntax of the form shown in section 4.3.1, which weaves aspect hooks into all functions and (non-builtin) class methods within the `mymodule` module object added to the local namespace of the importing stack³. Less formally: importing `with AspectHooks()` applies aspect hooks to all potential targets of advice in the `mymodule` package. The behaviour of Python’s `with` keyword is defined by more magic methods: any object with `__enter__()` and `__exit__()` defined can be used here, where `__enter__()` is run at the beginning of the enclosed block, and `__exit__()` when leaving the block.

PyDySoFu caches the original `builtins.__import__` object in an instance of the class, and replaces it with `AspectHooks.__import__`, in its `__enter__()` method. This is reversed by replacing `builtins.__import__()` with the cached object in its `__exit__()` function. `__enter__()` and `__exit__()` are the magic methods corresponding to entering and exiting Python’s `with` blocks, meaning that this technique modifies Python’s importing only when imports are made within a block such as that in section 4.3.1. The resulting implementation for weaving aspect hooks is uncomplicated, as can be seen in fig. 4.2.

³Python’s use of the stack namespace in its importing system means that careless re-importing a module can lead to multiple copies of it in different function stacks, meaning that the same name resolution (such as resolving a class by its name in a module) might, after applying aspect hooks in PyDySoFu, change the behaviour of procedures depending on where they are called. Scenarios where this might arise are deemed unlikely enough that the risk of this design decision becoming troublesome are considered negligible. Still, it would be remiss not to make note of the fact.

```

1 class AspectHooks:
2     def __enter__(self, *args, **kwargs):
3         self.old_import = __import__
4         import builtins
5         builtins.__import__ = self.__import__
6
7     def __import__(self, *args, **kwargs):
8         # ...replacement import logic for performing weaving...
9
10    def __exit__(self, *args, **kwargs):
11        builtins.__import__ = self.old_import

```

Figure 4.2: Magic methods used to enable the `with` keyword usage for PyDySoFu

4.3.2 Strengths and weaknesses of import hooks

As a technique for weaving aspect hooks, this new method provides multiple benefits. Application of aspect hooks is straightforward from the perspective of a modeller using PyDySoFu, whose code clearly applies aspect hooks and does so in a legible way for future maintainers. Import hooks’ explicit application of hooks to modules means that places where aspect hooks might be applied equally explicit, and thereby implements an interpretation of aspect-oriented programming’s principle of “obliviousness” moderated by clarity. While join points are oblivious to potential aspect application, callers of join points can expect to be more aware, and signs for codebase maintainers are left directly within source code. Aspect hooks can be applied to specific modules or every module depending on the use of the supplied `with` statement, allowing for total weaving or actual hook weaving [CS04] depending on their preferences. Further, performance is optimised in comparison to the previous implementation of PyDySoFu, as hooks are weave-able at a more granular level (on the level of procedures such as functions and methods, rather than all attributes of a class).

There are also caveats of this approach that are necessary to address. As aspect hooks are woven in the new implementation of PyDySoFu via Python’s import functionality, any procedure not imported from a module cannot have aspect hooks attached. **Consider adding local namespace weaving to pdsf3: should be easy to implement as a cheeky little monkey-patch...** However, as aspect orientation is primarily concerned with a separation-of-concerns approach to software architecture, targets are expected to exist in other modules, and we do not consider this to be a significant limitation.

A more significant limitation of the import hook approach is that the object with aspect hooks woven exists in the namespace of the module *importing* the join point. In other words, this method

makes it impossible for a module to make use of aspect hooks that are woven in an unrelated piece of code. *is this true??? verify, I think not, if imported normally. tested but didn't realise that pdsf won't fuzz functions with single-character IDs, so inconclusive.* We therefore have a “semi-oblivious” property to our aspect orientation approach: targets of advice are unaware of any adaptations made, but *any code making use of those adaptations must be aware enough to at least apply aspect hooks.*⁴

In a manner of speaking, this can be considered to alleviate some concerns with aspect orientation as a paradigm. Aspect Orientation is criticised for making reasoning about programs more difficult [Prz10; CSS04; Ste06]. One cause of this is that aspects separate logic from where it is run; Constantinides, Skotiniotis, and Stoerzer’s comparison with the jokingly proposed `come from` statement [Cla73; CSS04] is a reminder that it can be effectively impossible to understand how a program will execute if the path of execution is not at least linear or clearly decipherable from source code. Aspect orientation as a paradigm violates this linearity by design. However, import hooks as implemented in fig. 4.2 allow for aspect-oriented code which can be interpreted in one of only two ways:

- ① Looking at the original implementation of a procedure, its intended execution is clear. A programmer can make use of this directly and it is guaranteed to behave as expected.
- ② Any program making use of a procedure imported from a module will see, when the procedure is imported, whether it has had aspect hooks applied. In this case its behaviour is unknown — falling prey to the design flaws discussed in the aspect orientation literature ?????? — but this unpredictability is highlighted to the programmer.

As a result, while import hooks are somewhat limited in that they are applied specifically to imported code and break the traditional AOP concept of obliviousness in at least a weak manner, these two facts combine to arguably fix a latent issue in the design of the aspect oriented paradigm. The original PyDySoFu implementation was able to modify any procedure in a more traditional, oblivious manner. While this new implementation is clearly more limited as a result, we consider these limitations an overall benefit to the design of the tool, and a contribution to aspect orientation framework design.

⁴Note that once aspect hooks are applied, advice can still be supplied from anywhere in the codebase.

4.3.3 Weaving process

Weaving in PyDySoFu's updated implementation takes place via monkey-patching of aspect hooks, as described in ???. Aspect hooks replace executable targets within a module at the moment the module is imported. When the target is invoked, the wrapping aspect hook is executed in lieu of the original target object. The wrapping aspect contains the target function within its closure, allowing it to execute the original target; however, it was also created by the `AspectHooks` class, and so has reference to aspects registered against it.

The process of weaving with this implementation of PyDySoFu is therefore as follows:

- ① An aspect is imported using `AspectHooks`, as shown in section 4.3.1.
- ② An aspect is registered against `AspectHooks` as shown in fig. 4.3, by providing a regular expression matching the ID of any join points to apply an aspect to and the aspect to apply. Specifically relevant to aspect registration:

→ Methods permitting aspect application are:

- ① `AspectHooks.add_prelude(rule, aspect)`, which registers advice to be run before a target is invoked
- ② `AspectHooks.add_encore(rule, aspect)`, which registers advice to be run after a target is invoked
- ③ `AspectHooks.add_around(rule, aspect)`, which registers advice to wrap an aspect invocation, effectively providing the functionality of `prelude` and `encore` advice with a single aspect
- ④ `AspectHooks.add_error_handler(rule, aspect)`, which registers advice to catch and process exceptions raised by a join point
- ⑤ `AspectHooks.add_fuzzer(rule, aspect)`, which registers advice to modify a target before it is invoked, effectively providing aspect application within a join point, and permits arbitrary ast-level modifications of the target

→ Each method described above also accepts an optional `urgency` integer parameter, enabling optional additional features of PyDySoFu which are discussed in section 4.4.3.

- Invoking the above aspect hook registration methods compiles and caches the regular expression provided for efficiency when identifying matching join points in later invocations of aspect hooks.
- Each method outlined above returns a callback which de-registers the aspect provided, to facilitate the ephemeral application of aspects. This could be useful in experimental codebases where aspects represent behavioural deviations, and a researcher looked to run simulations with different deviations applied to compare datasets, for example.

A bit clunky
— take the
last sentence
out in favour
of notes in
discussion or
rewrite.

- ③ Invocations of executable objects within the module imported in ① are replaced with aspect hook wrappers of themselves; an invocation of an executable object within that module therefore triggers PyDySoFu's aspect hooks, which seek aspects registered against the ID of the target they wrap. Aspects registered against a join point identifying rule which matches the ID of the wrapped target can therefore be identified, and executed as appropriate.

All aspects applied when registering advice are callable objects; however, they have different function signatures. The arguments expected by each type of advice is:

prelude advice takes `target`, `*args`, `**kwargs`

encore advice takes `target`, `original_return_val`, `*args`, `**kwargs`

around advice takes `next_around`, `target`, `*args`, `**kwargs`

error_handler advice takes `target`, `handled_exception`, `*args`, `**kwargs`

fuzzer advice takes `ast_steps_from_target`, `*args`, `**kwargs`

...where `target` is the target an aspect is applied to; `original_return_value` is the value returned by a target after it was run; `next_around` is what an around-style advice runs when it has finished its pre-target component, and intends to run the target itself; `ast_steps_from_target` is a list of AST objects representing the definition of the target a fuzzing aspect is applied to; and `*args`, `**kwargs` is a Python idiom collecting arbitrary positional and keyword arguments passed to a function but not specified within its function signature, which are here employed to collect arguments

```

1 def log_invocations(target, *args, **kwargs):
2     '''
3     An aspect to print invocations of a target.
4     '''
5     print("Invoking " + target.__name__)
6     target(*args, **kwargs)
7     print("Invocation of " + target.__name__ + " finished.")
8
9 from pdsf import AspectHooks
10 with AspectHooks():
11     from some_module import some_func
12
13 AspectHooks.add_around("some_func", log_invocations)

```

Figure 4.3: Registering an aspect against AspectHooks

passed to a target, and can be used by any aspect to inspect the expected behaviour of the target (or to pass into the target, if the aspect invokes it).

The execution of `prelude`, `encore`, and `error_handler` aspects work predictably: `prelude` aspects are executed before the wrapped target, `encore` aspects after the wrapped target, and `error_handler` aspects within a `try` block encompassing the execution of all advice. Notably, the return values of `around` aspects replace those of their target, and `encore` aspects replace that of the target if they return any value which is not `None`. By comparison, the implementations of `around` and `fuzzer` aspect weavers are non-trivial.

As with all aspects, `around-style` aspects are given the target being invoked alongside its supplied positional and keyword arguments. However, these aspects are also given the next `around-style` aspect woven against the target to run. Each aspect should be provided its `next_around` argument. However: with a series of many `around-style` aspects to apply, and each calling the next in the series, a given aspect would need to know the `next_around` parameter for the aspect it calls; the second could call a third, meaning it would need to know the `next_around` parameter for its own successor, which could only come from the first in the series. The third could call a fourth, requiring another `next_around` parameter, which the second would have to pass to the third, and the first to the second. The inelegance of this naive solution to the problem seems confusing and inelegant for modellers to interact with or understand.

Decide whether the “-style” prefix I sometimes use should stay or go when copyediting. Or: when do I use it? Be consistent.

However, two alternatives present themselves. One is to use Python’s generator pattern, which can be simply described as a callable which returns successive items from a series on each invocation.

```

1 | from functools import partial, reduce
2 | final_around = lambda target, *args, **kwargs: target(*args, **kwargs)
3 | nest_around_call = lambda nested_around, next_around: partial(next_around
4 | target_with_around = reduce(nest_around_call, around_advice, final_around
    )

```

Figure 4.4: Simplified example of weaving around aspects in PyDySoFu

An alternative approach is to supply each around-style aspect with its first argument (the next in the series) before the aspects are invoked, unburdening individual aspects with responsibility to retrieve and apply the next aspect successfully. This approach is employed by PyDySoFu. The first argument of every around-style aspect is provided by constructing a partial function. Partial functions are a feature of many languages which allow for some arguments of a function to be provided, but for the function not to be invoked; instead, a new function is returned, with any remaining arguments left to be provided on invocation. With the first argument provided, the resulting partial function has the signature `target, *args, **kwargs`, where all values are known and trivially supplied by any around-style aspect.

The problem to solve is therefore the construction of the partially applied aspects in such a way that each refers to the next correctly. To solve this, each function is provided to its precursor; iterating through all around-style aspects in this manner results in a single aspect which contains a reference to the second, the second a reference to the third, and so on until the final registered aspect. The final aspect has no successor it can reference however; to overcome this, a function which does nothing but execute the original target is provided, with the same signature as the partially-applied aspects. The final function terminates the chain, and is responsible for executing the target itself; every other aspect is simply responsible for the execution of the next around-style aspect in its chain.⁵ In this way, around-style aspects can be woven in a manner which simplifies a developer's interaction with the aspects. A simplified snippet of the source code of PyDySoFu which implements this process is shown in fig. 4.4 to illustrate.

The final kind of advice to apply are fuzzer-style aspects, which PyDySoFu contributes to aspect orientation framework design. These aspects make modifications to the definition of the target they are applied to, allowing advice to be woven within a target. Arbitrary transformations of the target also

⁵Note that this is technically optional; if it determined it was necessary, any around-style aspect could break the chain by calling the target with its arguments directly.


```

1 | if fuzzers is not None and fuzzers != []:
2 |     code = dedent(inspect.getsource(t))
3 |     target_ast = ast.parse(code)
4 |     funcbody_steps = target_ast.body[0].body
5 |     for fuzzer in fuzzers:
6 |         non_inline_changed_steps = fuzzer(funcbody_steps, *args, **kwargs
7 |         )
8 |
9 |         if non_inline_changed_steps:
10 |             funcbody_steps = non_inline_changed_steps
11 |
12 |     target_ast.body[0].body = funcbody_steps
13 |     compiled_fuzzed_target = compile(target_ast, "<ast>", "exec")
14 |     if not isinstance(t, FunctionType):
15 |         t.__func__.__code__ = compiled_fuzzed_target.co_consts[0]
16 |     else:
17 |         t.__code__ = compiled_fuzzed_target.co_consts[0]

```

Figure 4.5: A code snippet representing the process of modifying a target of an aspect application using a fuzzer-style aspect. `t` refers to the target on which modifications are applied.

allow removal or change of any part of the target definition. This is achieved by acquiring the original source of the function to modify using Python’s built-in reflection library, `inspect`. The source can be parsed into an abstract syntax tree using Python’s built-in library, `ast`, which contains a list of AST objects representing a target’s original definition. This list of AST steps is then passed to a fuzzer aspect as an argument, which returns a new list of AST steps if any changes are to be made. The new steps are compiled and the resulting Python code object containing a modified function definition is monkey-patched into the original target, effectively changing its definition. The original code object is cached during aspect hook weaving, and is used to replace the modified code object after target execution to avoid unexpected behaviours in future invocations of the target. A simplified example of the process of acquiring, modifying, compiling, and monkey-patching a target’s underlying definition is found in fig. 4.5 .

4.4 Optimisations

Additional features were implemented in the presented incarnation of PyDySoFu which improve its usability in the research software engineering context, and optimise performance where necessary: deep hook weaving, non-dynamic weaving, and aspect priority support.

```
1 | from pdsf import AspectHooks  
2 | AspectHooks.deep_apply = True
```

Figure 4.6: Code snippet enabling deep hook weaving

```
1 | from pdsf import AspectHooks  
2 | AspectHooks.treat_rules_as_dynamic = True
```

Figure 4.7: Code snippet enabling dynamic weaving

4.4.1 Deep Hook Weaving

Under ordinary circumstances, PyDySoFu weaves aspect hooks into an imported module, but not modules imported *by* that module. Deep hook weaving enables weaving of aspect hooks into all potential join points by applying hooks to functions and methods imported by the original module, but also any recursive imports. A code snippet enabling the feature is presented in fig. 4.6.

Deep hook weaving introduces performance overhead due to additional checks for aspects which were woven dynamically. However, it is plausible that the modules within which desired join points reside are not imported directly by a developer, but are available to them indirectly through another package they make use of. It is also plausible that a developer may look to instrument the entirety of — and investigate details of — a call stack without attaching a debugger to a process. In both examples, PyDySoFu’s default behaviour is insufficient. Deep weaving is therefore provided to support these use cases.

4.4.2 Non-Dynamic Weaving

PyDySoFu’s runtime weaving of aspects allows for aspects to be applied at any time. Aspects can be removed and re-applied during program execution. However, programs may be written with the intention of applying aspects once — as program modifications to be introduced without direct manipulation of a codebase — so that, once defined, the set of applied aspects would remain static. In this scenario, the dynamic weaving of PyDySoFu introduces unnecessary overhead. Weaving aspects dynamically means that every invocation of a prospective join point requires searching for aspects which *presently* apply to it.

```
1 | from pdsf import AspectHooks
2 | AspectHooks.manage_ordering = True
```

Figure 4.8: Code snippet enabling priority ordering of aspect application

To avoid this overhead in scenarios where it is unrequired, PyDySoFu can cache the aspects applied when any callable wrapped by an aspect hook is invoked for the first time. On its first execution in this mode, the aspect hook stores the set of aspects it matched to the invoked target, and future invocations retrieve the set of aspects to apply from the cache. This avoids expensive regular expression matches, which fail in all cases but those where an invoked target is to be augmented by the application of an individual aspect the regular expression is paired with.

For performance reasons, the default mode of PyDySoFu is to run with non-dynamic weaving. Dynamic weaving functionality is enabled by toggling the `treat_rules_as_dynamic` flag on the `AspectHook` class, similarly to deep hook weaving (see section 4.4.1). A code snippet demonstrating this is shown in fig. 4.7.

Add metrics here on the performance implications; they're pretty significant, I actually added them because the regex matches slowed my own code down to a crawl. Our experiments make use of non-dynamic weaving for this reason.

4.4.3 Priority Ordering of Aspect Application

As dynamic weaving allows for non-deterministic application of aspects, it may be that the order in which aspects are intended to be applied is not the aspect in which they are woven. To support these use cases, aspects may be sorted by priority during the lookup of aspects to apply when a join point is invoked.

As mentioned in section 4.3.3, when aspects are registered against the `AspectHooks` class, an optional urgency parameter is available. The parameter takes an integer representation of priority of aspect application, with higher numbers representing more urgent application. Aspects with no urgency applied default to `urgency=0`. High-urgency aspects are applied before low-urgency aspects. This feature is disabled by default, but can be enabled using the code snippet shown in fig. 4.8.

4.5 Discussion

contributions
to what?

Is this list
complete?

The new implementation of PyDySoFu makes a few contributions , particularly in comparison to the previous version :

- Introduces a new technique of weaving aspect hooks on import, improving its design over a typical aspect orientation framework by making use of Python's `with` keyword when weaving hooks to trade a degree of obliviousness for clarity
- Aspect hooks can be applied with more precision than the previous implementation of PyDySoFu, meaning:
 - Users of the framework can better delineate between total and actual hook weaving
 - Unnecessary overheads from checking dynamically applied aspects at each join point are reduced.

PyDySoFu's current incarnation also provides opportunities for improvements and for future work. Our intended use case for aspect orientation for simulation & modelling is in scientific codebases specifically; direct integration with the scientific package ecosystem (which is vibrant in Python's community) should be made. A good initial project would be integration of aspect application in sciunit tests [Tha+17]. A discussion on potential use cases of PyDySoFu together with existing research software engineering technologies is provided in section 8.7 .

Chapter 5

RPGLite: A Mobile Game for Collecting Data

RPGLite is a game designed to mimic existing popular games, while being structured to permit exploring all game states via formal methods; Kavanagh and Miller have produced PRISM models which can be model-checked to identify optimal play strategies in all game states [KM20] . Some experiments were conducted around RPGLite to answer the question: “Over time, do players converge on an ideal strategy of play?” **Fix formatting and correct wording of William’s hypothesis. Review note from Jess: this is *very* important (and, yes! I don’t believe i have my RQs well-defined anywhere in the thesis yet, which is concerning.) Insert my own research question here**

Real-world play of the game invites many research questions. For example, an alternative question to answer would be, “what strategy of play do players typically adopt?” or, relatedly, “do all players adopt the same strategies?” ¹ Kavanagh and Miller’s work can identify the *cost* of an action [KM20; Kav21], allowing for richer datasets to be produced and more in-depth analysis of real-world play to be conducted. One possibility invited by these datasets is that it may permit modelling of real-world players’ interactions with the game, and therefore offers an opportunity to investigate the effectiveness of aspect-oriented simulation and modelling. Referring back to **TODO INSERT RESEARCH QUESTION HERE** , it would be feasible to build a model of RPGLite play which did not satisfactorily model real-world data, inviting aspectual augmentation of the model in an attempt to improve its performance, and thereby investigating the research question proposed.

¹These are not of research interest for the purposes of this thesis, but examples of interesting questions to ask of a game where “correct” and “incorrect” actions can be categorised.

About what,
and why?

Consider
changing
title. add a
sentence in
first para-
graph de-
scribing why
this fits in
with the rest
of the dis-
sertation,
and why it is
here.

Suggest new
sentence

PRISM unde-
fined

“game states”
not well de-
fined here,
though I do
define it later
on

Lift screen-
shots in this
chapter from
game-on pa-

However, it would not be possible to perform analyses of player behaviour without real-world player data. Datasets collected empirically would be required to compare against synthetically generated datasets, to ascertain their similarity and thereby compare different methods of producing synthetic datasets. As RPGLite is a real-world sociotechnical system which already yields a formal analysis of play, and datasets from this system can be used to support multiple avenues of research in different disciplines,² producing a dataset of real-world RPGLite could also support future work in a variety of fields as a non-commercial dataset for research in fields including at least game design and gameplay research [Kav21], formal methods [KM20], sociotechnical simulation research, and research software engineering tooling demonstration.

J: rework
passive voice

To that end, a collaboration was undertaken with Kavanagh to develop and release a mobile implementation of RPGLite which would collect player data for research purposes. The dataset produced would enable Kavanagh to demonstrate the utility of their model checking in an empirical scenario [Kav21], and for the purposes of this thesis, it would also enable the analysis of models representing player behaviour, by supporting the comparison of these models against the collected data.

J: this para's
easier to un-
derstand
than previ-
ous ones in
this chap-
ter. When
reworking,
consider how
the others
could be
rewritten in
a style closer
to this? She
also notes it
might belong
further up in
the chap.

The development and release of RPGLite as a mobile game offers an opportunity to make use of aspect orientation in a new context: a model of naive RPGLite play can be produced which represents random play, and aspects could be written which augment the naive model representing hypothesised player behaviour. If the aspectually augmented models generate data which correlates with empirically sourced data more closely than that generated by the naive model (with random play, and no aspectual augmentation), we can dismiss naive play as “realistic”, and understand the aspectually augmented behaviour as “more realistic”. Many aspects can be written representing different styles of play, which might be adopted by different players, a concrete benefit of aspect orientation in modelling & simulation.

The modelling & simulation research opportunities RPGLite presents are realised in chapter 6 and chapter 7; this chapter focuses only on the design and development of RPGLite as a mobile game, and the dataset collected from real-world play and released [Wal+20; Wal+21]. It starts with an overview of RPGLite itself in section 5.1, and progresses to a discussion of the game's implementation in section 5.2

J: avoid pas-
sive voice

²Later contributions in this thesis are supported by datasets produced by RPGLite, as were the contributions of Kavanagh's PhD thesis [Kav21].

J: She sug-
gests moving
this earlier,
too

before concluding with a discussion around the empirical data collected using the game in section 5.3.

5.1 An Overview of RPGLite

Needs copy-
editing

RPGLite is a simple two-player game played in turns. Each player selects characters independent of the other, with each character having a unique set of abilities and properties, which are generally health, chance of success on attack, and damage dealt on a successful attack. The abilities of some characters necessitate additional properties. Each player selects an “alive” character (one with health greater than 0) to perform their action against a chosen “alive” target (or occasionally targets). A successful attack — randomly determined by chance of successful attack for the selected attacking character — results in that character’s unique ability being inflicted on their target[s]. A random player is chosen to take a first move, players may always skip their turn as a valid action, and players continue to take alternating turns until a victor is left with the only “alive” characters.

Eight characters are available for selection, with the following abilities :

Knight Deals damage to an opponent character on a successful hit.

Archer Deals damage to two opponent characters on a successful hit.

Wizard Deals damage to an opponent character on a successful hit, disabling (or “*stunning*”) them for the duration of the opponent’s next turn.

Healer Deals damage to an opponent character on a successful hit, and heals themselves or, optionally, the other player character instead (assuming that character is still alive).

Barbarian Deals damage to an opponent character on a successful hit, dealing additional damage if their health is low when attacking.

Rogue Deals damage to an opponent character on a successful hit, dealing additional damage if the target’s health is low when attacked.

Monk Deals damage to an opponent character on a successful hit, and immediately takes another turn, until their attack is unsuccessful.

J: compress
the descrip-
tion list.
Maybe make
it bulleted?
I wonder
whether it
can just be a
para or some-
thing.

J: whose?
the barbar-
ian’s? Clar-
ify.

Gunner Deals damage to an opponent regardless of success, dealing additional damage on a successful hit.

Specific details of each character — their health, chance to hit, and damage on hit as well as character-specific details (such as the threshold for additional Barbarian or Rogue damage, for instance) — are defined as a “*configuration*” of RPGLite. Different configurations change the game’s “*balance*”, a term referring to the relative strengths of different characters or character pairs. For example, if a configuration leaves many characters with initial health values close to a Barbarian’s threshold for additional damage, then they become a very powerful character due to their ability to inflict additional damage. If the Monk’s chance to hit is high, the repeated turns it offers can be very advantageous. Character skills can work in concert with each other: choosing a Barbarian and Healer such that the barbarian can be kept at low health for additional damage, but the healer can be used to keep them alive, may be an effective strategy depending on the game’s configuration. Kavanagh et al. found that model-checking a configuration of the game could discover the relative strengths of characters and character pairs when played optimally [Kav+19].

RPGLite’s design has two objectives it must meet. First, that it is interesting to players, which requires that it is approachable and complex enough not to be immediately solvable. This is necessary for real-world data collection, and to demonstrate a design representative of something that could conceivably be a real-world game with an active playerbase. Second, RPGLite’s design must be sufficiently simple for model-checking. Model checking is a necessary requirement of design because of our need to identify optimal moves: analysis of player behaviour rests on our understanding of how close to “ideal” players are, and whether players approach ideal strategies over time. This is the crux of the work found in formal methods research on RPGLite [Kav21; KM20], which relies on a reduced state space in order to calculate optimal moves and character pairings.

5.1.1 RPGLite’s Design Implications

The state space of RPGLite makes it unusually well-suited to analysis through formal methods. Because of this, datasets produced through simulation of RPGLite can be compared against two other datasets: one of real-world play, and another of what can be mathematically shown to be “correct”

Jess noted this wasn’t formally described. I describe it later, but clearly it’s confus-

player behaviour.

To demonstrate this state space, note that RPGLite games can have their states described by a set of values: the healths of characters on each team, plus a stunned character. With eight characters each having their own maximum health value, which we denote w, x, y, z ³, two players, and an indicator of which character is stunned.⁴ The entropy of a game's state is therefore $\log_2(w \times x \times y \times z \times 3 \times 2)$, where the multiplications by 3 and 2 represent the stunnedness indicator for the current player (either character, or neither), and the player whose turn it is to play, respectively. The maximum health for a character is 10 hit points, which makes the maximum entropy of a game state $\log_2(10^4 \times 3 \times 2) = \log_2(60000) \approx 16.87\text{bits}$. Each player picks 2 of 10 characters, and can choose to attack either opponent character with either of their own, or skip their turn, for a maximum of 6 possible actions. We can therefore see that the total entropy of the entire RPGLite game is no greater than $\log_2(10^4 \times 6 \times 2 \times \binom{10}{2}) \times 6 \approx 24.36\text{bits}$.⁵

Iterating through these states allows us to map the entire state space of RPGLite. As the state space the game defines is relatively small⁶ it is feasible to analyse every possible game state. Note that the figure of $\approx 24.36\text{bits}$ includes movements *between* states as well as the states themselves, meaning that it is feasible to map the potential progressions through all possible games of RPGLite using formal methods. Further, this allows us to understand the chances of a player winning given transitions between different states, for example by representing moves in the game as transitions in a decision diagram where nodes are the game's state. We can calculate exactly how "good" a move is, by comparing chances of success when making a given move in a given state against the chances of success when making the objectively optimal move.

In this way, RPGLite's design allows it to be understood formally, yet it also draws on common

³Maximum health values are dependent on RPGLite's configuration.

⁴Note that stunnedness is valid for exactly one character for one turn, meaning that only one character may be stunned at any time, and the status effect immediately resets, meaning there are only three possible states for stunnedness: either character belonging to the player taking a turn, or neither.

⁵The actual figure is smaller:

- Players may choose two characters with special abilities that prevent them from attacking both opponent characters at once (this accounts for 9 out of 10 characters), giving 5 possible actions in a turn, rather than 6.
- The "metagame" **Metagame is never defined!**, which refers to the perceived "best" strategies at any given point in time, would impact the chance of a player selecting making certain moves or choosing to play with certain characters. RPGLite is designed to be slightly "unbalanced" in the parameters of different characters such as health, attack damage, or potency of special abilities, meaning long-term players are expected to learn effective playstyles and adjust accordingly.

Players' behaviour is therefore less uniform than this calculation would imply, but the calculation provides a higher bound on the entropy of the game.

⁶For comparison, mapping valid positions in chess takes about 136 bits [Nie77], and this figure does not account for valid moves within the game, which our calculation for RPGLite does

J: so is that the state, or just a description of a subset? (It's a complete and uniquely identifying state, but I don't make this clear enough)

J: define earlier? Also: why is entropy important here?

J: does the order matter?

J: Summarising her notes but 1. Footnote should be integrated into body and 2. Entropy's a confusing concept, wasn't what she was expecting here for defining a state space. Either make the concept extremely clear or, better, avoid the concept of

game design elements and is sufficiently “interesting” to generate data from a real-world playerbase.⁷

J: first section isn't clear, she notes she didn't understand. In second section she circles “known”, asking, “known by whom?”

This yields some useful properties for the purposes of aspect-oriented simulation :

- ① Simulated moves can be selected naively, i.e. at random, but can also be made perfectly according to the known-correct move in a given game state, or made with some calculated “cost” as to the chance of winning.
- ② Real-world players’ behaviour can also be analysed according to the same metrics: for example, moves made by players can be analysed to understand bias, whether players learned to play “better” moves over time, or whether they selected known-strong characters more frequently than those who can be formally shown to have a relatively low chance of winning games.
- ③ As the actions taken when playing RPGLite are consistent — such as deciding the target character of an attack, or a character to use in an attack — random play can be simulated as a “naive” play style, which can be compared against real-world players. Where player behaviour does not correlate to naive play,⁸ the biases of players may be represented as aspects which are applied only to specific actors within the simulation.
- ④ Should aspects be suitable as a manner of accurately representing biased play, aspects offer a separation of concerns within the simulation: any nuance found within the playstyle of specific real-world players would be replicated and applied not to the model itself, but to specific simulated players. Playstyles might also be mixed with the application of multiple aspects.

Footnote suggests correlation is defined in chap 6; include a more specific cref.

Whether aspect orientation is suitable for the realistic simulation of RPGLite gameplay is the topic of the remainder of this thesis. However, the design of RPGLite allows for a controlled system where a clear notion of “good” and “naive” behaviours can be defined, the system is closed insofar as all interactions within the system are known and all game elements are precisely understood, and all interactions take place between experiment participants for data collection purposes, allowing for a large dataset to be collected without information being removed due to players not consenting to their data being collected and disseminated for science. In short, RPGLite’s design constitutes a system where all aspects are well-understood, no interference is anticipated from system components which are unknown or outside of experimental control, and lots of data can be collected for analysis. Therefore, the data generated by

⁷Demonstrated through data collected from several thousand completed games, which can be found at [KWM20].

⁸The concept of play correlation is introduced in chapter 6.

gameplay is suitable for comparison against datasets outputted through aspect-oriented simulation. This can be used to assess whether simulated players with aspect-affected behaviour accurately reflect the playerbase, and so can help to assess whether aspect orientation is suitable for realistic simulation.

5.2 Implementation of RPGLite

J: compress

Knowing ideal play is useful. However, to understand how real-world players would interact with RPGLite, empirical data needed to be collected. To produce this, a mobile online multiplayer version of the game was developed for data collection purposes. Play constituted engagement with an experiment for data collection, and after several months, a database logging player behaviour offered a dataset which could be used to simulate real-world player behaviour.

This section describes details of RPGLite's implementation as a data collection tool. Some lessons learned after reflection on the implementation process were documented for the benefit of others' avoiding our errors[Wal+21] .

5.2.1 Consent to participation

Players gave consent to participate in a scientific study as a part of creating an account to play RPGLite. Players were required to explicitly scroll through an information sheet and consent agreement and agree to both to make an account. Copies of both were made available to download for player reference in both the game and RPGLite's website⁹ . Players were instructed to contact the involved researchers to withdraw participation, and were instructed that they could do so at any time. Email addresses to contact were noted in both the app and the website.

The study, information sheet, consent form, and data collected were reviewed by the University of Glasgow Science & Engineering ethics committee before the game was publicly released.

⁹RPGLite's website is available at <https://rpglite.app/>, where copies of the consent form and information sheet are linked.

Should a summary of the paper be a section here, or maybe an appendix? Keen not to plagiarise, but it's also maybe worth including. William and I both spent a lot of time getting that paper out. I presented it at GameOn2020. He's included some in his

5.2.2 Mobile app

As a mobile game, RPGLite’s user-facing component was an application, distributed through the Google Play Store on Android and the Apple App Store on iOS. This was developed in Unity, a framework for developing games in C# which can be distributed to almost any platform. Most assets were developed in GIMP, with character designs contributed by a commissioned artist online. Unity allowed for a “WYSIWYG” or what-you-see-is-what-you-get interface builder, with event handlers defined in C# code which would “hook” into events signalled by interface element interactions.

J: rest of this para should be made shorter.

User-facing components of the game were largely produced by Kavanagh, the collaborator on this project and original RPGLite designer. Therefore, in an attempt not to take credit for this component of the work involved in collecting the RPGLite dataset, curious readers are referred to Kavanagh’s notes on the development process for full details [Kav21].

Beta testing required user engagement. Apps were deployed to Android and iOS devices of colleagues, who played a series of games to check that game logic was sufficiently robust and graphic design sufficiently adequate for public distribution of the game. In a manner inspired by design science methodology [JP14], beta tests comprised of iterations on the artefact of the game and its server-side component (discussed in section 5.2.3) which were distributed to beta testers. Informal feedback and enhancement requests were sought on each iteration until the game behaved correctly in all edge cases (no major bugs were reported) and a final design was settled upon (no major complaints about interaction or aesthetics were reported), at which point the game satisfied criteria for public distribution.

An example of the game’s visual evolution is given in fig. 5.1, where colourful buttons replace a tabular, text-heavy interface. Another example is the evolution of the game’s main screen, “Active Games”, which loaded when players logged into the app and allowed users to see and interact with active games they were involved in. The visual identity and colour palette of this screen was refined iteratively, as can be seen in fig. 5.2. Other features which were developed following user feedback while beta testing include the implementation of a notification system and a leaderboard showing a player’s experience relative to their peers; these are shown in fig. 5.3.

There are some good screenshots in the paper for RPGLite;

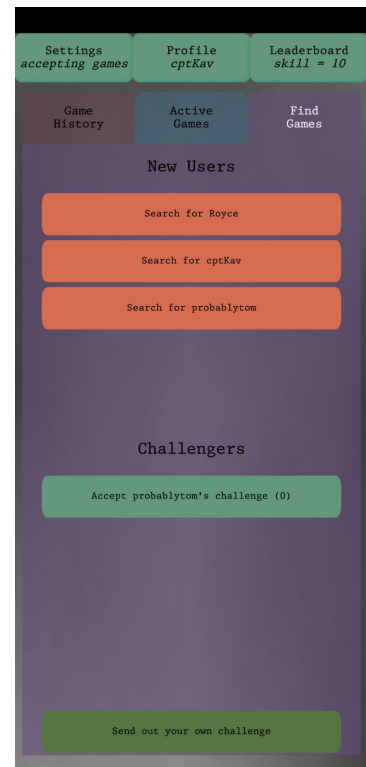
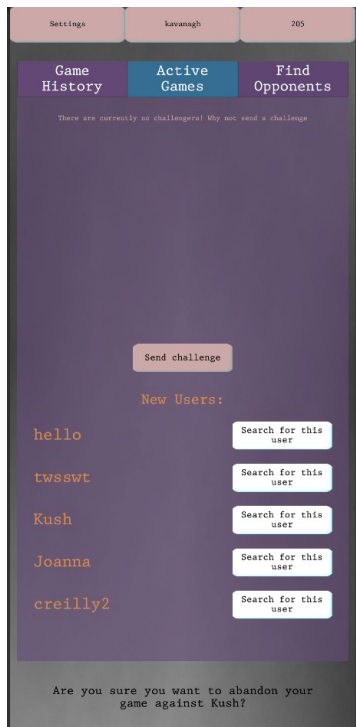


Figure 5.1: Evolution of the “Find users” matchmaking screen. Prototype left, final design right.

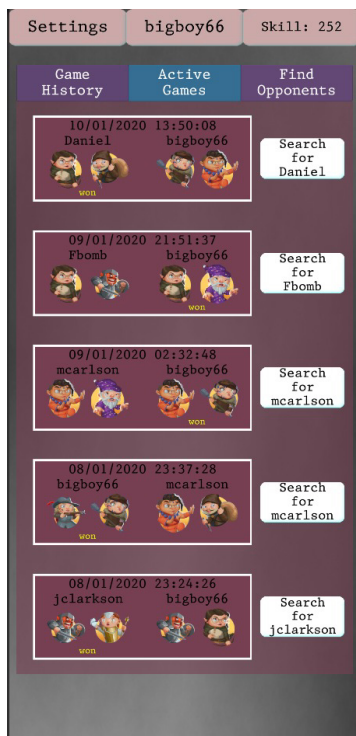


Figure 5.2: Evolution of the “Active Games” screen, which allowed RPGLite users to see and interact with games they were playing. An early prototype is shown to the left; a refinement through beta testing in the center; and the final version released to the public to the right, with an improved colour palette.

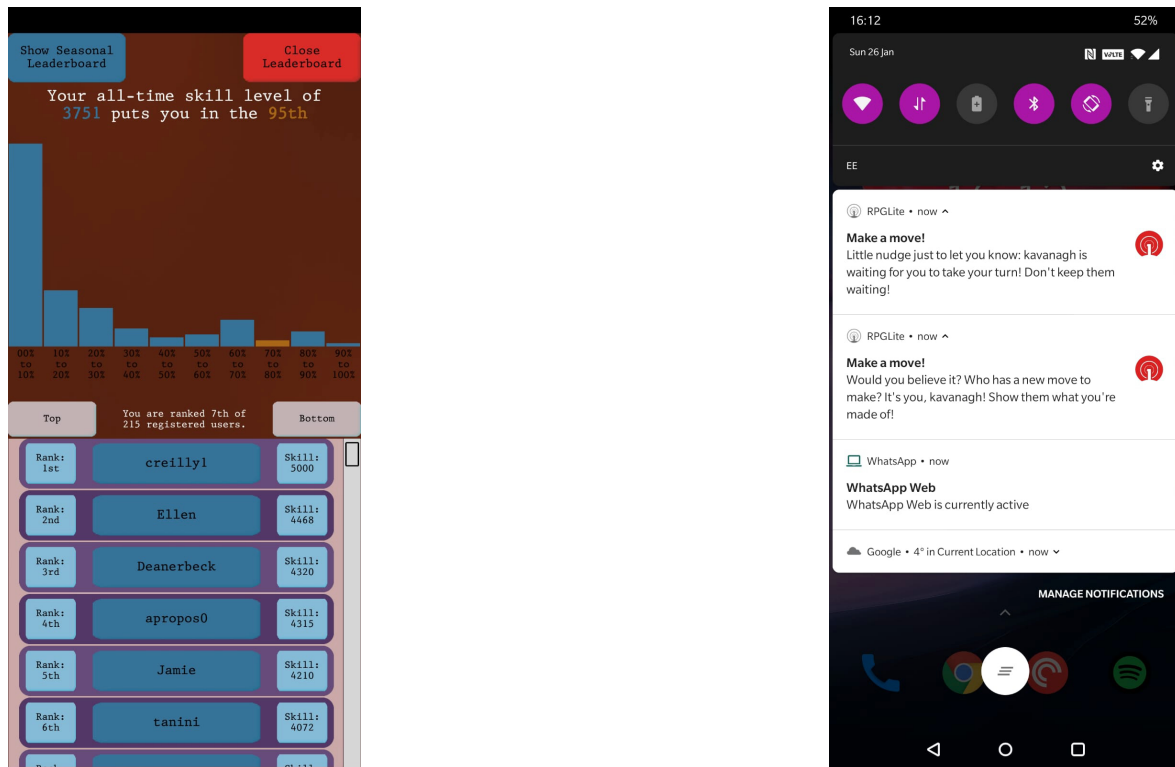


Figure 5.3: Features developed following requests from beta testers. A leaderboard of player engagement is shown to the left; examples of notifications when an opponent has acted are shown to the right.

5.2.3 API & Server-side Logic

J: how could the data not be empirical?

As the data collected ought to be empirical, RPGLite was developed as an online game. This required a server and API for a client to communicate with.

A REST API was developed with Python's *Flask* framework. Endpoints were created to support all in-game actions requiring centralised state (or a recording of that state), allowing for player search, matchmaking, player profile design, game history and statistics analysis, ranking calculations, login and password reset, synchronisation of access to sensitive information, and other in-game activities. The API also allowed moves to be made, and rejected erroneous game states or unauthorised input from any malicious actor, to ensure that the data collected, analysed, and published was not manipulated. The API also sent push notifications to an opponent's device when moves were made. This was a feature requested during beta testing, which was anecdotally observed to increase engagement with the game.

On each of these actions, data was collected about the action performed, and logged in a database. In addition, in-game activities which required no server-side input but were considered to have potential

in later analysis would log datapoints through the API into the same database.

A MongoDB database instance was installed and managed on a University of Glasgow Computing Science virtual machine. The no-SQL nature of the database permitted flexible structuring of the data, and simplified analysis of the games' results after data collection was completed. The API was also hosted on the same virtual machine. A combination of port access rules applied via Nginx and hardening the security of the database itself through MongoDB's account features and controlled permissions within the server prevented unauthorised access to the database, ensuring that the data remained untampered.

Game state was interpreted client-side. The intention in this design choice was to un-burden a centralised service we were responsible for maintaining, moving functionality such as calculating game states after moves were created to clients on players' devices. This left the serverside logic mostly concerned with executing database transactions and ensuring the integrity of game states was uncompromised. Reflections on the suitability of this approach made after the game's launch have been published separately [GameOn2020].

5.3 Empirical Play and Data Collected

In total, players produced a dataset of 9,693 games. It also includes 1,105,065 datapoints generated by gameplay or player interaction with the client, such as players checking their position on a leaderboard, searching their game history, or rolling to hit as part of an attack. The data is drawn from the MongoDB database used to run the game. With player consent on signup (see section 5.2.1) this data is published and available for all researchers to analyse in the format of a JSON object [KWM20].

J: suggests this is separate from games. Is that correct? (Yes; make this clearer.)

Completed games drawn from the MongoDB instance contain many fields. The breadth of datapoints logged by player interaction with the RPGLite app means there are an extremely broad range of datapoints available to curious researchers. A full list of available data is available in a text document within the dataset ?? ; in brief, some of particular interest include:

J: make more concise. (I made slight edits when adding this revnote, maybe the para can be

→ The history of moves made, and the times those moves were made

- The players involved and the winning player of each game (by username as no personally identifying data was collected)

ELO unde-
fined.

- The ELO scores of players in each game
- The characters chosen by each player

J: “score”
is vague. I
clarify in the
footnote but
it should be
in the body
here. Also,
“score” is not
of interest.
We don’t
use it in the
experiments,
don’t waste
attention on
it.

- The “score” of each player ¹⁰

RPGLite was played through two “seasons”: after an initial 3,000 games played in a first configuration of RPGLite, a second configuration was released. This meant that, having learned a strategy to play with the initial configuration (a preference of characters, for example), changes were made which may invalidate what players had learned. The dataset therefore contains two conceptual subsets: data collected from a system with the same mechanics but slightly tweaked J: this chapter is actually unfinished. (I can’t believe I missed this. OK. There’s a little more to say on the dataset itself; set yourself up in particular for the parts of the data that are important in the experiments. Make note of the changes between S1 and S2; I think I’d meant to look this up, forgot, and thought the final section was complete, which is a serious oversight, but there’s not too much left to write here.)

¹⁰RPGLite’s mobile app presented users with a naive scoring mechanism used to rank users on a leaderboard which some users then used to identify other players of a similar notional skill.

Chapter 6

Simulation Optimisation with Aspect

Orientation

My main concern with this chapter at time of writing (10-10-23) is that it goes into too much trivial detail (e.g. types passed into different functions), and not enough into substantive detail (experimental design and why things are set up the way they are). Maybe I have enough substantive detail, but just a surplus of trivial detail? Consider this carefully when redrafting and kill your darlings. If you're not Tom and you're reviewing this, please be generous in suggesting things to eliminate!

This chapter describes the design, implementation, and results of an experiment which investigates the use of advice to augment modelled behaviour. The experiment's foundation is RPGLite, described in chapter 5, which was developed to collect data on real-world play of a game with formally verifiable correct moves [KWM20] [Cite William's thesis here somehow?](#) . This dataset yields further research opportunities: data produced by simulated RPGLite play can be compared against the real-world dataset to assess whether a simulation is accurate. The experiment presented in this chapter makes use of this opportunity to answer the research question: **REFERENCE THE RESEARCH QUESTIONS BEING ANSWERED, THEN BRIEFLY SUMMARISE THE EXPERIMENTS BEING CONDUCTED, WHY THEY ANSWER THE QUESTION[S], AND FINALLY GIVE A RUNDOWN OF THE STRUCTURE OF THIS CHAPTER.**

6.1 Aims

Aspect orientation's use in previous simulation and modelling efforts have typically focused on the use of aspects to compose model or simulation details **There must be tons of good citations for aspects being used to compose together a simulation / model** . Critics of aspect orientation note that the act of process composition makes visually understanding codebases difficult, and so ensuring that a simulation properly models real-world behaviour is made trickier with the introduction of aspect orientation. However, aspect orientation might instead be used to *augment* an existing model, by rethinking what aspects are used to represent.

An alternative use of aspects would be to first build a non-aspect-oriented model of *expected* behaviour, and separately build aspects which describe deviations from this. For example, one might more realistically simulate safety procedures by first producing an idealised, “naive” model of what employees are expected to do, and separately model alterations to prescribed behaviour as an employee's boredom, expectation that checks and balances are unnecessary wastes of their time, and so on — effectively, separating out models of degraded modes[JS07].

Previous research on the use of aspect orientation to model degraded modes adopted the traditionally claimed benefit of aspect orientation: separation of cross-cutting concerns, allowing for a greater reusability of codebases[WS18a]. A repository of cross-cutting concerns in socio-technical simulation such as boredom was developed as a library to be applied to any future models[SW16]. However, aspects used in simulation have no intrinsic need to represent concerns that are cross-cutting. Indeed, whether they can be accurately used to represent cross-cutting concerns in simulation is the topic addressed in **Add a cross-reference to the chapter on cross-cutting concern simulation accuracy when it exists** . Aspects might instead be used to represent *amendments to processes* which deviate from an expected norm, in this case represented by the idealised model aspects are applied to.

To more concretely relate this to the experiment at hand: play of RPGLite can be modelled as players matchmaking, picking characters, and then mutually taking turns until one player's characters are entirely expired. Once a player's characters are dead, new matches can be made. This can continue indefinitely. Lacking a heuristic to select next moves or characters, players might be modelled as picking random moves. However, heuristics for move selection can be added to the naive model of

We don't think this is the case — those citations just don't exist! Rework to discuss use of AOP in business process modelling, and note that nobody made those models in practice, just tooling.

play by way of augmenting the processes already defined through aspects. This approach can be of significant utility in both modelling player behaviour and accurately modelling different players:

- ① Different players might use their own unique heuristics to model play. Each player's behaviour is therefore well described by separating what play "looks like" to what makes a given player play differently to their peers.
- ② Different players might lean more heavily on different heuristics, or mixes thereof. Play might be characterised by reliance on experience, on recent games, on knowledge of an opponent, and so on; these different variables can be expected to be weighted differently by each player, adding complexity to the code which models this individualised play.
- ③ A modeller might discover a new idea for a heuristic long after developing an original concept for a model. Due to the impact of ②, ideal architectures for an approach such as this should allow these heuristics to be defined entirely separately to the base model to maximise flexibility when maintaining experiments.

Considering ①, ②, and ③, architectures and paradigms which enable separation of concerns are well-suited to defining alternative approaches to play. Some architectural approaches such as mixins or plugin design patterns might support this structure well, but they typically rely on language features (in the case of mixins) or knowledge of software engineering (in the case of design patterns). Aspect orientation is typically provided to developers as a framework or runtime in a language (such as AspectJ[Kic+01] or PROSE[PGA02]) and can require minimal architectural understanding to use: concepts are simple, and the effort of composition is alleviated by the supporting framework or runtime.

The approach makes little use of aspect orientation's significant contribution — cross-cutting concerns — as whether behaviour cross-cuts different parts of a codebase is not of interest in this use case. Instead, aspect orientation is treated as a composition mechanism with a reasonably low degree of technical knowledge required.

6.1.1 PyDySoFu Suitability

Some aspect orientation frameworks do not adequately achieve this requirement. For example, the most influential framework, AspectJ, requires the use of language extensions to define integrate aspect orientation[Hil+00], and similar additional complexity is added in seemingly every alternative framework, through the use of bespoke virtual machines, compilers, translators, or languages[rajan2006nu_towardsAO_invocation; PAG03; SL07; BH02b].

PyDySoFu, however, requires very little additional knowledge to use. Its design prioritises simplicity and a shallow learning curve that makes its adoption by researchers without a software engineering background feasible: **maybe cut this list of reasons PyDySoFu is fantastic...**

- PyDySoFu is implemented as a pure-python library, meaning that it can be installed through Python’s package manager (pip) and imported like any other Python library. No additional supporting infrastructure is required.
- Aspects in PyDySoFu are simple functions which take as arguments whichever pieces of information are pertinent for the function’s use as an aspect¹.
- To weave a PyDySoFu aspect requires only a method call, which returns a **callable** which unweaves that aspect.
- Defining PyDySoFu pointcuts requires only a regular expression matching a method name. This can apply to a wide range of join points if required, but where method names are provided directly, the join point is made clear.
- Additional clarity over where aspects *can* be woven is introduced by PyDySoFu’s transparent weaving of aspect hooks, mitigating some of aspect orientation’s most prominent criticisms.

PyDySoFu therefore satisfies the requirements of this work well: it offers composition of procedures outside of the scope of an original codebase, makes what is being composed where clear to a programmer, and makes no significant changes to Python as a language (thereby requiring users to specialise in fewer tools).

¹For example, an “encore” aspect which is woven after a target procedure returns will be provided that target’s return value.

6.1.2 Proposed Experiment

Aspect orientation's use as a composition tool for model components makes sense in principle, but it is unclear whether the addition of behaviours to a naive model would make the model more "realistic". Furthermore, changes to a model could alter its representation so as to weaken its mimicry of the system it simulates; adding behaviours could make it *less* realistic. The fundamental issue at play is that it is unclear whether the changes made would properly represent what might be empirically observed. While PyDySoFu's design makes understanding what is being composed simpler than other aspect orientation frameworks, a composed model under this paradigm is still split across multiple areas of a codebase, making a visual assessment of whether a model accurately reflects the intended behaviour impractical. It is therefore important to demonstrate the efficacy of PyDySoFu and the modelling paradigm it introduces, by confirming the realism of a model to which behavioural variation is applied.

We can confirm whether aspects can realistically represent changes to a naive understanding of the real world by comparing their output against empirical data. For example, if a such a model of behaviour in a system outputs data which correlates poorly against empirically collected data, a change to that system would make it more realistic if it improved this correlation, and could be said to be realistic if the generated data appeared sufficiently "close" to the empirical dataset — which here means that the correlation between the two is of statistical significance. Such a change can be aspect-oriented. Therefore, we can see the application of aspects as the application of packages of potential improvements to a base model, which can be verified by way of comparison to known-good datasets.

This is the basis of the experiments in this thesis.

With datasets collected empirically on RPGLite's play, we can build a naive model of play and aspects to apply that should realistically model data from players. This can be used to answer the research questions motivated in section 2.4.1 **THESE RESEARCH QUESTIONS HAVE BEEN AMENDED SINCE I FIRST WROTE THIS SECTION (I hadn't decided on a wording then) AND THE PLACEHOLDER SUGGESTION THAT PREVIOUSLY EXISTED WAS JUST WHETHER ASPECTUALLY AUGMENTED MODELS COULD BE REALISTIC. THIS CHAP DOES MORE, SO THIS SECTION NEEDS TO BE REWORKED. :**

“

- ① *Can changes to a model be represented as advice using aspect orientation?*
- ② *Can a model be made more realistic by applying aspect-oriented improvements?*
- ③ *Can advice which improves the realism of a model be ported to others without loss of realism?*

”

To answer these questions, a naive model of play is produced. Aspects are developed which model learning within the system defined by the naive system. The synthetic datasets produced by models with naive or aspect-applied models can be compared to an empirical dataset sourced from real-world players of the game, and their similarity compared. **Rework to include details on the other RQs, too .**

“Naive” is used here to describe a model which does not encode any understanding of the players of the game being modelled. Traits such as learning, distraction, or aptitude for similar games are irrelevant to the naive model. We need a naive model to demonstrate the effectiveness of an aspect-oriented alternative: we can measure how closely it reflects empirical data, and compare this against the same measurement drawn from another model with behavioural variations encoded using aspects. A closer match from our aspect-oriented model would demonstrate that the technique can enhance a model’s realism. Our naive model also provides a null hypothesis: if no improved similarity is observed, the technique brought no improvement to the model’s realism. No measurable difference indicates that weaving behavioural variations as aspects has no impact on a model’s realism.

Write a little here on the learning aspects.

Flesh this out as a brief wrap-up of our experimental technique. The end of our “naive” model explanation might be useful to move down here / rework into this para. Contrasting the similarity of the empirical dataset to both naive and aspect-applied datasets... This discussion is provided so as to provide context for the following sections; experimental design is discussed in more depth in section 6.5.

6.2 Naive Model

A naive model of play was developed by separating each stage of the actions taken by players in the client-side app, and separating them into individual procedures.

To facilitate the retrieval of information pertinent to a simulation in an applied aspect, the model was written so as to contain simulation state as mutable function arguments. The model was written as a workflow, and state of workflow execution was separated into three components: the actor that a function invocation (or “step”) represents activity from; the context of that step in the execution of a workflow; and the context of that workflow’s execution in a broader environment. Incidentally, we found this structure to allow a flexible and natural implementation of a procedural simulation, which should translate easily to existing simulation frameworks such as SimPy[Mat08]:

Actor — allows the function to identify the actor performing the activity defined by the function.

This argument is any object uniquely identifying an actor.

Context — allows the function to determine details of the current thread of work being undertaken by the actor. This is necessary because in some simulations, the same actor might pause and resume multiple occurrences of the same activity — for example, they might concurrently play three different matches in RPLite. As a result, it is necessary to understand the context of the action being performed by the actor in question. This argument can be any object uniquely identifying the context of a piece of work, but should be mutable (such as a class or dictionary-like object) to permit the communication of information across invocations of different action-representing functions.

Environment — an actor’s actions are often determined by the global environment they act within.

There may be ancillary details to the actor’s actions and the context of their particular thread of work which they are undertaken within which are used to determine behaviour, such as a landscape they traverse or other actors they might choose to interact with. Because all actors share access to a global environment, this also provides a message passing space, or a space where actors can set values and flags other actors might look to, should those details be more general than their specific thread of work at a given point in time. ²

²This is different to environments in some other simulation frameworks, such as SimPy[dev21], where the environment

Each simulation step receives these three arguments at a minimum. Because steps of the model are functions, and therefore valid join points, aspects applied to these have access to the entire state of the simulation.

The naive model of RPGLite follows a simple workflow mimicking player interaction with the client-side application used by real-world players. A graphical representation is provided in fig. 6.1. **Position this figure properly on the page. Maybe reword? Pretty verbose.** The model it describes produces synthetic data. It does so by simulating players interacting with each other in a broader RPGLite ecosystem. The large number of players is important: future behavioural variations introduce behaviours which guide future actions based on past experiences. Non-determinism early in the model can therefore cause different players to act in different ways. Simulating many players allows the system to represent a wide variety of early experiences, as would occur to real-world players, too. As the aspects being woven simulate learning, early experiences which cause players to learn maladaptive strategies — or simply strategies less optimal than those learned by their peers — should later contrast against winning strategies, so that players with less useful early inferences can learn from players who play more optimally.

Consider changing fig. 6.1 to only include details about the model of the game itself, leaving the steps about data analysis to the later aspect-applied version (which is the real experiment). Move fig. 6.1 to .svg for fidelity The stages of the naive model as laid out map to those encountered by real-world players. Two randomly-selected players repeatedly select characters to play (from the pool of 8 characters available in the real-world game), and a player is chosen to play first at random. That player selects a random valid move to make.³ The active player alternates, and the process repeats, until such time as an active player starts their turn with both of their characters fully depleted of health. The player with remaining characters is the victor, and another game is started by picking random players and starting a game between them, until a predetermined number of games has been played.

After a sufficient number of games are played, analysis of the various datasets collected can begin, although this is discussed in more detail in section 6.5.

controls scheduling and execution: this structure imposes no constraints such as models of time, and anticipates that any such functionality should be implemented by the programmer. However, an environment such as SimPy's might satisfy a programmer's needs when using this particular pattern.

³Many simulated decisions are random; this is because the model is designed to be naive, so it avoids informed decisions where possible. Informed decisions are expected to be woven later as aspects.

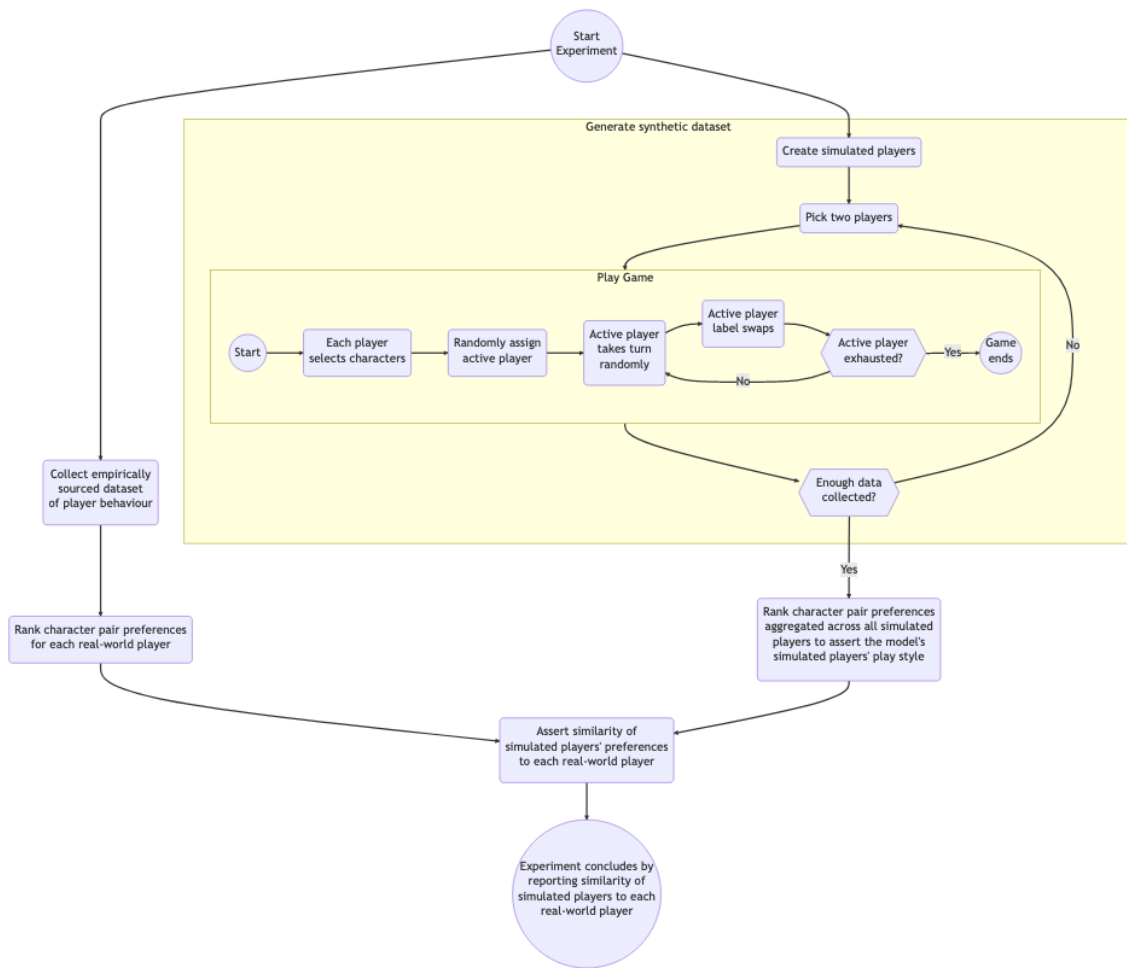


Figure 6.1: A flowchart diagramming a simple model of RPGLite play, without aspects applied.

6.3 Experimental Design

The naive model above describes a simulation of RPGLite itself; however, the models exists to support investigation into research questions, and cannot do this alone. To answer our research questions, the model must be augmented so as to represent real-world play. This section describes the modifications made to the naive model (through aspects applied by PyDySoFu) and how data from those simulations can be analysed to arrive at answers to our research questions.

The questions these simulations seek to answer are:

- ① Can we fit model details per-player to get realistic player behaviour?
- ② Can we cluster based on accuracy of different models for each player?
- ③ Can we generate predictive data for unknown models from known ones?

These RQs are taken from an early outline of a thesis structure. Maybe they need rewording?

To investigate these, we produce models of learning which can be applied to the naive model to augment player behaviour. Simulated players will have their behaviour augmented in relation to the character pairs they choose to play. Real-world players could reasonably be expected to choose character pairs which are stronger together over time. In addition, it is possible to calculate objectively optimal character pair selections in RPGLite [Kav21]; therefore, players should be expected to gravitate toward more optimal character pairs and away from poorer choices as they become more experienced in the game.

A few facets of gameplay are expected to be learned by players over time, and would theoretically also be suitable objects of simulated players' learning. One analog to character pair selection could be move selection: players would be expected to make increasingly optimal moves as they play. However, data can be muddled by players understanding the optimal use of some characters' moves better than others, as different characters invite different styles of play. Small changes to game states (the opposing team's composition, amount of health remaining, opponent play styles) could influence player behaviour also. Fundamentally, the state space of moves to be made and scenarios in which to make those moves is extremely large, making simulation computationally expensive. By comparison,

character selection happens before gameplay starts, eliminating in-game factors which might influence player decisions.

Modelling the learning of character pair selection allows for modelling of players learning the metagame, rather than the game itself. Modelling learning of the metagame allows us to model players learning about the game and other players' interactions with it, rather than learning the game itself. An example of the distinction is the difference between learning how optimally make moves with a Barbarian in a team, and learning the mistakes other players commonly make when playing a Barbarian to exploit their weakness. As RPGLite is "solvable" — there is a provably correct way to play the game — players picking sub-optimal character pairs can be expected to observe that they are frequently beaten by better ones, increasingly favour pairings they identify as successful, and so converge on strong character pairs. Convergence en-masse would indicate a stable metagame. This was observed in practice in real-world RPGLite play [KM21]. However, while players made fewer mistakes over time when playing RPGLite, the average cost of a move to their probability of winning increased after many games (though decreased after a few). [KM21]. Player behaviour which converges on provably optimal play is simpler to represent than players' initial improvement followed by a seeming loss of performance. RPGLite Character selection is therefore a more convenient metric to model players' interaction with than move selection, as it presents a smaller state space for simulated players to explore while also providing a simpler expected end state for those simulated players to converge to.

Models of learning are therefore applied to players' character selections. Our primary goal in applying models of learning is to demonstrate whether players can be accurately modelled by augmenting a naive model to represent them in particular, thereby answering the first research question. Different real-world players are expected to exhibit different styles of learning; thus, multiple models of learning will be applied. Different players might also learn differently, but in the same style; for example, two players might exhibit similar biases, but one could be quicker to learn from experience or another more cautious in the application of new knowledge.

To identify models of learning and parameters for those models which most realistically represent real-world players, we look to optimise the parameters for each model, for each player, by running multiple simulations of players learning in a particular manner (with particular parameters), identifying their preferred character pairings, and calculating the similarity of the simulated character pairings

to that of a player in the real world. By doing so, we can anneal to a parameter which represents the optimally realistic model of a player learning in a given style. If the data produced by this optimally parameterised model of learning is not similar to empirically sourced data with statistical significance, we determine that the model of learning applied does not represent how a player learned in the real world: even the closest dataset the model produced fit its target dataset poorly. However, if a parameter can be found for which the model reliably produces character pair preferences which align with the real-world players' with statistical significance, then the model can be said to realistically represent that player's learning of the character pair metagame.

This section will first discuss models of learning generally in section 6.3.1, and will explain how aspects implementing these models can be produced in section 6.4. The strategy used to implement these models of learning is discussed in section 6.3.2. The design of experiments investigating the aforementioned research questions by applying these aspects is then described in section 6.5. **Rearrange either this paragraph of the sections it refers to so they're in order / make sense. At time of writing they refer to 6.3.1, then 6.4, then 6.3.2, then 6.5...!**

6.3.1 Models of Learning

Different people learn in different ways. Indeed, no universally-accepted definition of learning appears to exist. This is presumably because it is convenient to define what it means to learn differently in the context of different pieces of work. Cognitive models of learning can be useful when considering mental processes specifically, for example, whereas functional models of learning could lend a more empirically applicable perspective. What it means to learn is clearly outwith the scope of this thesis; however, our experiments will include models of learning. To justify our model, we consider a functional approach to learning, as considering learning in this way appears more closely linked to the empirically focused work of modelling real-world behaviour than alternatives.

Lachman identify[Lac97] that standard definitions of learning summarised as, "[...] a relatively permanent change in behavior as a result of practice or experience" have practical shortcomings such as a focus on behavioural change (as learning may not change behaviour) or conflating learning's process and its product (the process by which we learn is not obviously identical to its result, of which

behavioural change is an example). They suggest learning might be better defined as:

“ (...)the process by which a relatively stable modification in stimulus-response relations is developed as a consequence of functional environmental interaction via the senses [...] rather than as a consequence of mere biological growth and development [Lac97].”

(...)

They note that their definition distinguishes learning from phenomena such as injury, changes to one's maturity, or sensory adaptation, incorporates stimulus-response relationships the research community consider as learned, and differentiates learning's process and product. Their model is inherently functional, making it useful for the purposes of simulation and modelling, although they offer only a definition of learning and a brief comparison to the standard textbook definition they introduce. The work presented is not intended to demonstrate its improved model of learning empirically, only to discuss its semantic merit. However, the models proposed in this thesis require only a theoretically informed, sound basis for their model of learning, and a lack of empirical justification is not a barrier to the relevance of the model Lachman have proposed.

De Houwer, Barnes-Holmes, and Moors propose a functional definition of learning which is primarily concerned with providing a definition of learning which is both accurate and useful for the purposes of cognitive learning research [DBM13]. Doing so attempts to provide a model around which some consensus can be reached; learning is a central concept in psychology, and they see their definition as supportive of cognitive work. They introduce their definition as follows:

“Our definition consists of three components: (1) changes in the behavior of the organism, (2) a regularity in the environment of the organism, and (3) a causal relation between the regularity in the environment and the changes in behavior of the organism.”

This model of learning contains more nuance than the “textbook definitions” of learning they paraphrase as, “a change in behavior that is due to experience”, but does not stray far from the core concept: some environmental stimulus impacts behaviour in a causal fashion. Their introduction of “regularity” to their definition refers to the presence of the stimulus with some form of repetition,

whether this be multiple instances of a stimulus at different times, or the same stimulus occurring concurrently. De Houwer, Barnes-Holmes, and Moors explain that such a model is straightforward without the sweeping inclusivity of the simple model mentioned earlier, and is easily verified (although, as in the work of Lachman [Lac97], empirical verification is omitted in favour of semantic analysis).

Aside from other benefits more particular to their research community, these benefits are especially useful from the perspective of modelling learning in our case. A simple, functional definition can be captured in a software model, and introduces few opportunities for misunderstanding or misapplication. It also introduces helpful concepts — such as regularity and causality — which the other definitions discussed do not. For the purposes of this thesis, we therefore adopt this definition as a basis for our model of learning.

6.3.2 Modelling Learning in RPGLite

We use De Houwer, Barnes-Holmes, and Moors’s definition of learning [DBM13] to arrive at a model of learning which is encodable in aspects that can be applied to our naive model of RPGLite. We are interested in modelling players learning a preference for character pairs over time. The model should therefore account for: how player behaviour is influenced in accordance with their learning, in line with the definition’s first criterion; repetition of experience in successive games influencing the direction of a player’s learning, in accordance with the definition’s second criterion; and do so in a causal manner, in accordance with the definition’s third criterion. We therefore look to model a causal relationship between a player’s observation of successful character pairs and their future choices of character pairs.

To fulfil these requirements, a model might draw on previously successful character pairs to determine future ones. There are many ways in which this could be done. For example, we could model learning as consistently playing the character pair which most recently was observed to win a game. Any game ending naturally determines a winning pair, and we can select this pair when playing future games, until a different pair is observed to win instead. However, this does not align with our expectations around how players *would* engage with a game in the real world. Having a strategy one is confident in and being unlucky with the game’s random nature is unlikely to deter a player

from what they believe is ideal. Indeed, we can expect players to understand that perfect play might not be winning play: in some games, the right moves might not lead to a successful outcome due to moves randomly missing opponent characters. Equally, players may take time to become confident in a strategy; we would expect a player to explore character choices before settling on a preferred pair early in their experience, and would expect very experienced players to choose characters based on what they have learned, rather than continuing to explore options for which they have sufficient information to reason about. We can infer that:

- There are scenarios where players can be expected to observe wins/losses without incurring behavioural change.
- Players' confidence in what they have learned can affect their inclination to draw on what they learn when making decisions.
- What players learn in successive games would have a small impact in their early experiences, but an increasingly significant impact proportional to their experience in the game.

The model of learning used to simulate players' improved play of RPGLite can be explained following a similar structure: a component pertaining to observations players make about winning characters; a component pertaining to their inclination to use that knowledge when choosing characters (rather than exploring their options); and a component pertaining to the strength of that inclination in proportion to their experience.

The first — that players should make observations about winning players — can be implemented by modelling the choice as a probability mass function (PMF). The PMF maps character pairs to their chance of being selected by a player, and initially tracks all character pairs as having an equal chance of being selected. After every game, the chance of selecting the winning character pair increases, and the chance of selecting any other pair decreases. The sum of probabilities of being selected across all character pairs always sums to 1 (100%). This can be implemented as a record of the winning character pairs observed by a player: many ways of producing a PMF from a sequence of wins exist, but for the purposes of explanation, one such method is to take the proportion of wins for every character pair as their probability of being selected. This method produces a valid PMF because the probabilities must sum to 1 by virtue of the winning pairs being the sole determinant of the probabilities; every win

contributing to the probability of a character pair being chosen means that 100% of character pairs are considered, and therefore the probabilities inferred from that data also sum to 1.

A model of learning where prior belief as to optimal character pair is causally related to those which were previously successful fulfils the first criterion of the model of learning proposed by Lachman [citelachman1997learning](#); however, players would be expected to explore a state space in the early stages of their experience. The model proposed by Lachman identifies that the experience of a learning agent draws from “regularity” in their environment; we therefore require that an experience affects behaviour not when it originally surfaces, but after repeated exposure to it. This fulfils the intuitive understanding of learning discussed earlier — players’ confidence in what they have learned affecting their inclination to act based on experience — our second criterion. We therefore require some model of confidence an agent has in what they have learned.

The third inference — that confidence should be low initially, and grow proportionally to experience — explains the “shape” of the confidence model. We require a monotonically increasing function mapping experience, quantifiable as games played, to confidence, being a percentage chance between 0% and 100% that a player is to determine their character choice based on what was learned, rather than on an exploration of their space of possible choices.

Such a model of confidence offers a model of learning when combined with a PMF representing historical observations of a character pair’s success. If the model of confidence indicates that the player is to explore their space of possible choices, they select character pairs randomly; if they are modelled as confident enough that their behaviour is instead informed by their experience of their environment, they select a character pair according to the PMF their historical observations define. As the PMF affords higher probabilities to repeatedly winning character pairs, player behaviour is causally affected by the regularity of their experience, implementing a realistic model of learning in agreement with the functional model proposed by Lachman [[Lac97](#)].

Our model of learning is therefore complete, but a model of confidence remains to be defined.

6.3.3 Modelling Player Confidence in RPGLite

To model learning, we must find some function mapping experience (quantified by a count of games played) to some probability between 0 and 1, and fitting the criteria described in section 6.3.2.

A sigmoid curve fulfils both the second and third inferences in section 6.3.2. It also notionally conforms to expectations around “confidence” as an anthropomorphic trait: much like a sigmoid, confidence starts low, and remains so until it achieves some inflection point, after which it grows rapidly until tapering asymptotically toward some maximum. However, not all real-world players might express the same traits in their growth of confidence; the shape of their sigmoidal confidence may differ. In order to account for this while guaranteeing a monotonically increasing mapping of the number of games played to a player’s confidence, we select a sigmoid which can be parameterised to alter their shape.

A sigmoid curve is suitable for this model of learning where other curves would not suffice, because we require a period where players lack confidence and explore their options. Sigmoid curves such as the logistic [Ver45] or Gompertz [Gom15] are widely used when modelling systems [WJ97], but while they fulfil the role of a monotonically increasing curve with asymptotically low and high initial and final states, the shape of such curves is not trivially modified to fit different players’ learning styles.

More flexible asymptotic curves were developed by RICHARDS [RIC59] drawing on growth curves developed by Von Bertalanffy [Von38], which afford a natural pattern of growth; RICHARDS amends this curve to offer a parameterised growth rate. This curve can be made equivalent to other curves, including the logistic and Gompertz [FT+84]. This curve allows for a parameterised rate of growth, but lacks parameters controlling the points at which growth occurs most rapidly. The relative rate of confidence gain is a separate concern to the point at which such growth occurs: a player might cautiously grow in their confidence until they are already very experienced, or might bullishly grow in confidence yet plateau early, taking longer to reach complete confidence in themselves than they did to garner an initial increase, regardless of their relative growth in confidence.

The flexibility of a parameterised relative growth rate appeals to the notion that different players would gain confidence at different rates, but the point at which confidence accelerate most must also be controlled. We therefore employ the Birch curve, proposed by Birch [Bir99] for its increased flexibility

as compared to the Richards curve combined with its additional parameter used to control the curve's shape.⁴ Different players might exhibit different rates of growth in their confidence, and might grow maximally in their confidence at different points in their experience; the birch curve satisfies these properties as a model of confidence. **Investigate: did we ever model confidence with anything other than $c==1$, i.e. a standard logistic curve?**

6.4 Aspects Applied

I don't believe I make any note in this section that selecting character pairs based on a known distribution from player data isn't really learning. It's useful, though, because it allows us to demonstrate that using aspects can make us *really* accurate. So, it answers our research question, kind of, in that the aspectually augmented model is more accurate than the naive one. However, that aspect isn't particularly portable to a new system (we'd still be selecting based on the S1 distribution, not S2, so we'd expect it can't be ported) and doesn't so much change simulated behaviour as patch a part of the model. It's still useful to build models of learning, see which players are accurately represented by different models, and see whether those models are portable. I still have to figure out how to "sell" this properly...

Having a naive model of player interaction in RPGLite allows for the generation of a control dataset, which can be compared to aspect-applied datasets to examine whether the aspect-applied models are provably "closer" to empirically-sourced data than the naive dataset⁵. To generate the experimental dataset, the naive model producing our control is augmented through the application of aspects. These aspects fulfil different functions, and can broadly be grouped into three categories:

- ① Aspects implementing behavioural models, changing the behaviour of simulated players;
- ② Aspects instrumenting the naive model to perform observations necessary for the implementation of behavioural models (such as models of learning) on the naive model;
- ③ Aspects altering the behaviour of modelled players to simplify experimental observation, and to

⁴Birch refers to shape to mean the point of inflection of a curve. The point of inflection of an exponential rise to a limit is at its initial point; the point of inflection of the logistic curve is in the exact midpoint of the curve's growth.

⁵"Closer" here refers to a similarity measurement which we will define later in this chapter.

handle exceptions introduced by changes within the aspect-applied model.

The first set of aspects implement the behavioural change we anticipate will produce datasets closer in similarity to empirically sourced data than that sourced by the naive model. The second set lays the foundation for applying these changes. The third applies behavioural changes which experimental observations require. Necessary changes unrelated to our models of learning will first be explained in section 6.4.1, followed by an introduction of foundational observations for our models of learning in section 6.4.3. Finally, the models of learning built on those foundations are introduced in section 6.4.2. **The ordering of this explanation of different uses of aspects is confusing; once the aspects are explained individually, reorder the sections, rework the introduction to the structure of this section, and ensure the outline of the section reflects the changes made.**

A diagram of a game of RPGLite with all possible aspects applied⁶ is presented in fig. 6.2, by way of high-level overview of the aspects this section introduces.

6.4.1 Aspects for model improvement

Ensuring the Best Move is Played

As discussed in section 6.3, the goal within this experiment is to build a model of players' learning of character selection rather than move selection. However, randomly selected moves are liable to place players in unrealistically weak positions, as players are unlikely to make obviously poor moves such as skipping a turn with no clear reason. It would be unexpected to see true randomness in players' moves. This is a concern for the modelling of players' character selection, because the model of learning defined requires a causal relationship between what is observed (in this case characters which most reliably win games) and behavioural change (players choosing character pairs proportionally to their estimated chance of winning a game); if selecting random moves cause simulated players to lose games when they would have realistically won them, this would affect character selection by definition.

Character selection and move selection are therefore somewhat linked. However, move selection is most frequently optimal: in the majority of cases real-world players chose the best move available to

⁶Including aspects which should not be woven in the same experimental run: some aspects implement different models of learning and are therefore conceptually incompatible.

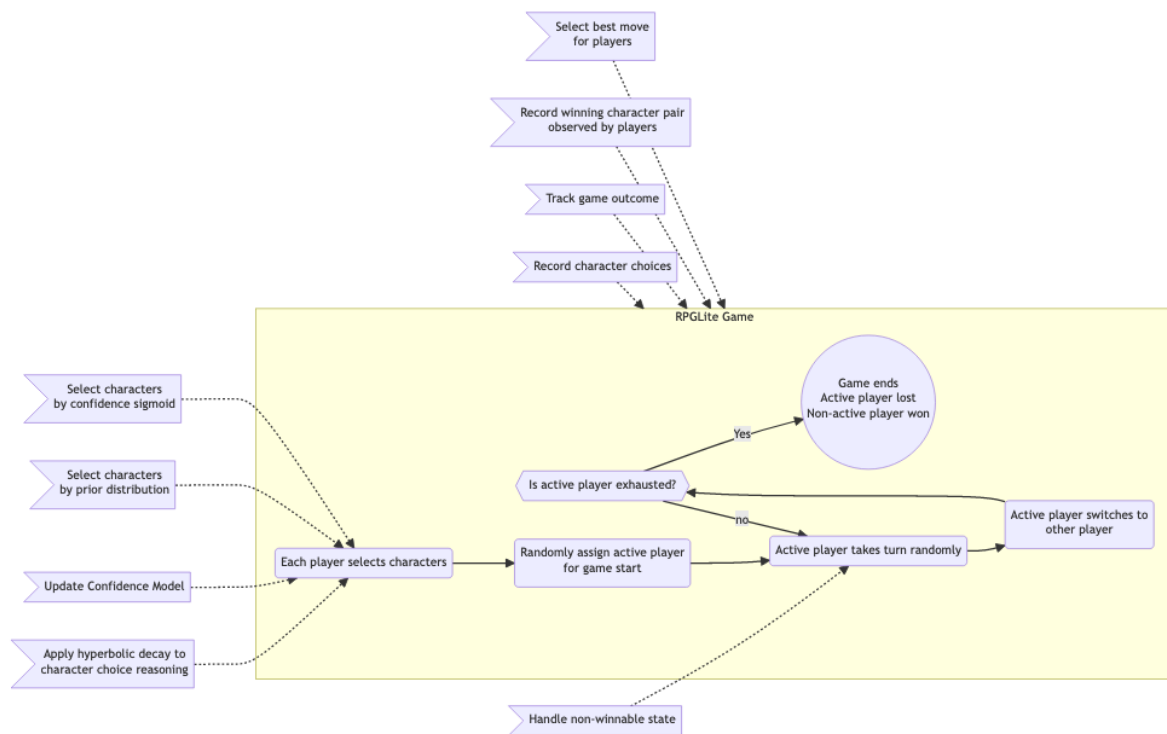


Figure 6.2: A flowchart describing a simulated game of RPGLite, and all aspects woven into the game to implement the various models of learning. Some aspects should not be woven together in the same experimental run, as they implement different models of learning.

them, as analysed by Kavanagh and Miller [KM21] :

“ (...)the majority of actions taken were optimal, with a cost of 0.0. In total 73% of the player’s season 1 actions and 77.8% of their season 2 actions were optimal.^a”

^aDifferent seasons of RPGLite are discussed in more depth in [Make a reference to the chapter\(s\) where I explain different seasons of RPGLite.](#)

(...)

Therefore, a naive model of move selection which represents real-world move choices in $\frac{3}{4}$ of cases is to select the best move at every opportunity. We therefore apply an aspect to the function responsible for move selection which performs a lookup on the dataset of action costs defined by Kavanagh and Miller [KM21], and select the known-optimal move in every case.

One additional factor in move selection is handled by this aspect. As simulated players’ behaviours are augmented to select universally optimal moves, an anomaly in the dataset was identified. In some games, optimal play would result in infinite loops: players would arrive in a state where they would skip turns mutually and indefinitely. This is because it is possible that, in RPGLite, to make any move is disadvantageous.⁷ Both players can concurrently exist in this state. Therefore, both players’ optimal move is to skip their turn. Kavanagh and Miller note that such states were reached in 64 cases in real-world play, but real-world players never skipped their turn; therefore, this scenario is also an example where suboptimal play can be re-introduced. This aspect therefore returns random moves if the previous two moves in a game were to skip. Such a state indicates that the game reached a stable point where neither player would optimally deal damage to the other. Random moves destabilise this state and return to ordinary play.

A model of move selection more closely aligned with expectations of player behaviour is beyond the scope of this thesis: a demonstration that models of both move selection and character selection can be optimised concurrently to arrive at a realistic model of empirical play with multiple aspectually applied models of behavioural variance invites itself as future work. [Do I need code snippets for notes on aspects?](#)

⁷This is because barbarians deal more damage once they lose a sufficient number of hit points. As a result, to deal damage to a barbarian can result in their having an opportunity to win in one move rather than two in certain health states.

Handle Game States with no Viable Moves

A consequence of playing games using only optimal moves was that unexpected states could occur. One might assume that, choosing moves optimally from a dataset of available moves in every game state it would be infeasible to arrive at states not present in this dataset. However: the dataset was produced through model checking, with the aim of identifying the cost of an action regards its impact on a player's chance of winning. If a loss is guaranteed, all moves have a 0% chance of winning; the model checker producing the dataset of available moves therefore identifies that the game is a foregone conclusion, and contains no valid moves.

In such a situation, identifying the winning player is as simple as identifying the player who made the previous move. We therefore apply an aspect which handles exceptions raised by move selection (as lookups will fail due to the game's current move state not existing in the dataset of action costs) and, after identifying that the exception indicates that the state does not exist (in this case, a `KeyError`), the aspect handles the exception by assigning the losing player's characters 0 health and swapping active players so that the simulated game proceeds with the losing player taking the following turn. As this turn starts with the active player having 0 health, the game ends as expected.

6.4.2 Aspects for Instrumentation

As explained in section 6.4, the experiment being discussed requires observational apparatus to be implemented in order to collect data about the experiment for later analysis. Ordinary experimental setups might require this apparatus to be tangled through the experimental codebase; however, aspect orientation provides a technique for augmenting the simulation to make observations before, during, or after any step of within the model. This demonstrates observational scaffolding as a cross-cutting concern as suggested by Gulyás and Kozsik [GK99]. Aspects realising this concept and supplying observations and additional models in support of our model of learning are discussed here.

Update Model of Confidence

The models of learning described in section 6.4.3 have players make observations as to the character pairs which most frequently win games, and have those players alter their behaviour in response to these observations. However, as discussed in section 6.3.3, a model of confidence is required to complete our model of learning. This model of confidence is a sigmoid curve, and the Birch curve was selected for our model of confidence. The role of this aspect is to update the birch curve when a game ends.

An aspect updating our model of confidence is applied to act after a game ends (an “encore” in PyDySoFu’s parlance), and updates a value in the model’s environment for confidence models. Confidence is represented as a mapping of player objects to floating point values between 0 and 1. The aspect supports modelling confidence in many ways, identified by a `sigmoid_type` parameter within the model’s environment.

- ① A logistic curve

Record Prior Distribution of Character Preferences

Keeping this in because I might need to explain the aspect, but it’s only used in the implementation of choosing based on prior distribution, which I don’t believe is ever applied. So! Decide whether to get rid of the writeup of that model. If I *am*, then get rid of this too. If not, add it to the learning models writeup below and explain it here.

Record Character Pair Choices

It is important to keep track of the characters chosen by simulated players, as this data is required to analyse the outcome of the experiment. Do we describe an experiment singular or experiments plural in chapter 6? outlined in this chapter. However, the naive model makes no special accommodation for the collection of character choices made by players. Character choices could be calculated by iterating through the list of completed games after a simulation is complete, filtering for games relating to a specific player. However, the collection of character selection data can be simplified by tracking what

This is unfinished!
There’s an inline todo note commented out in the source below that I’ve kept to remind me of what I was thinking when I left this in the state it’s currently in.

was played at the end of a game. This is achieved with with an aspect which observes the character pairs played in a completed game, and appends each players' chosen pairs to their own list. This aspect is applied after its join point executes (an "encore" in PyDySoFu parlance [reference pdsf's explanatory chapter / re-engineering chapter here? unsure which is most appropriate / introduces the concepts first.](#)).

There are many ways to collect this data, aspectually or otherwise; however, the simplicity of collecting information mid-process without requiring modification of a base model demonstrates the flexibility of an aspectual augmentation of models and simulations as an approach. Collection of chosen character pairs mid-simulation is an opportune example of this convenience.

Track Detailed Outcomes of Games

Similarly to the recording of character pair choices, the outcomes of games must be tracked. This information is important as an input of our model of learning, as successes and failures observed by players for each character pair can inform their models of learning. Some models of learning implemented, such as one including a hyperbolic decay bias (to be explained later, in section 6.4.3), require not only knowledge of the proportion of wins each character pair has achieved, but also the history of wins and losses: players might be biased toward recent win/loss observations, discounting older experiences in favour of the new. Detailed historical win/loss information must therefore be recorded.

As with character pair choices, the naive model was not engineered with the intention of providing this information specifically. However, the model is easily instrumented to collect such information at many suitable points. As with recording a player's chosen character pairs, an "encore" aspect was implemented which records wins and losses observed for character pairs on game end. Specifically, the aspect models players observing character pairs which won at the end of a game, lost at the end of a game, and also pairs which were uninvolved in the game. This requires a list of outcomes (wins, losses, and neither) for every player, for every character pair. The lists of observed outcomes for a given character pair record `True` for a winning pair, `False` for a losing pair, and `None` for a character which was not involved in the game's outcome. Each player observes the relevant state for every character

at the end of any game they play. Provisions were made in the implementation of this aspect for a player only paying attention to their own outcomes (i.e. only recording whether their own pair won or lost, without consideration for their opponent's outcome), though in practice all simulations were performed with players observing both their own outcomes and those of their opponents.

Record Winning Pair observed by Players on Game End

Unlike the detailed game outcome data recorded by the aspect described in section 6.4.2 and used in advanced learning models such as that of hyperbolic decay (to be explained later, in section 6.4.3), other learning models required only simple outcome data. The simpler model of learning making use of a distribution of winning teams to be explained in section 6.4.3 is more easily implemented against a simple list of winning teams. Rather than coalescing the more complex dataset collected as described in section 6.4.2 to achieve the desired format, another aspect can be applied to collect the data in a simpler format directly.

An aspect was implemented which also models players observing winning character pairs on game end, but collects the data in a simple list of character pairs observed to win games, rather than a mapping of character pairs to game end states as described in section 6.4.2. Character pairs which lose games or are not involved are not recorded for the purpose of the alternative model of learning.

6.4.3 Aspects Implementing Models of Learning

The aspects described above implement setup for the models of learning applied in this thesis, both instrumenting the underlying RPGLite model for observation and tweaking behaviours of the naive model. Two aspects remain unexplained, implementing the learning models themselves.

The first model of learning left to be described is implemented as described in section 6.3.1, drawing from a record of character pairs observed to win in earlier games to select characters in future games, in accordance with a model of confidence which activates either this behaviour or a randomly selected pair. A detailed description follows in section 6.4.3.

The second model of learning left to be described implements a similar logic, but introduces a

weighting on the history of winning pairs a player observes, introducing a hyperbolic discounting bias to historical observations. In this model, players increasingly discount old observations in favour of recent ones. A model of learning with a hyperbolic discounting bias is described in section 6.5.

Why didn't we make a bayesian model of learning? Should we have? Would this be difficult at all?

Character Selection using confidence sigmoid

The first model of learning applied through aspect orientation draws on character pairs a simulated player observed to have won games (as recorded by the relevant instrumenting aspect, discussed in section 6.4.2). The record of previously winning character pairs defines a probability mass function by selecting a character with equal probability to their rate of appearance in the history of winning characters.

Selecting a character based on this distribution is gated by a confidence model as described in section 6.3.3. If a simulated player's confidence model indicates insufficient experience to found decisions on, players will instead select characters randomly, effectively exploring their space of possible choices. By doing so, players have the time and opportunity to observe many matchups between different character pairs. Time to observe matchups is important because some character pairs may only present a strong choice if played against specific alternatives. We can imagine a character pair which is extremely effective against 50% of pairs, but extremely likely to lose games played against the remaining 50% of possible opponents. Without exploring possible matchups, a player may lack observations which would inform them about whether a character pair is effective in general, or in a narrow set of circumstances. A confidence model encouraging early exploration promotes the experience of a wide variety of matchups, avoiding this issue.

Another concern early exploration mitigates is that simulated players need opportunities to build well-rounded priors: if a character pair is never selected by any player, experienced players basing their character pair choices on those which previously win games cannot include pairs which were never selected, because they will never have had the opportunity to win games at all.

$$\text{score}(D) = \frac{1}{1 + kD}$$

Figure 6.3: Hyperbolic discounting score calculation

This aspect was implemented as an “around” aspect in PyDySoFu’s parlance, allowing it to apply additional logic before and after its join point. It could equally appropriately have been implemented as an “encore”, applying logic only after its join point had executed; the aspect discards the random selection made by the base model unless the simulated player lacks confidence. Either implementation can return a different character pair choice to the join point’s caller, allowing the chosen character pair to effectively be overridden, and so allowing the aspect-applied RPGLite model to proceed executing as it would otherwise; the only difference being the pair of characters “chosen” by the simulated player whose behaviour was augmented.

Character Selection exhibiting hyperbolic decay

Character selection with hyperbolic decay works similarly to the simple model of learning as described in section 6.4.3. Character pairs are selected randomly if a player lacks confidence or, if the confidence model applied to the player indicates that character pairs are to be chosen according to their historical observations of winning pairs, their record of winning pairs in previous games informs their choices instead. However, this model also applies a hyperbolic discounting bias which weights the player’s historical observations when making choices. *Maybe we should have also produced models of other discounting mechanisms, like exponential discounting (which I believe is actually the logically sound way to discount). Would be interesting to see whether some players are well-modelled by other discounting mechanisms. Investigate this!*

Hyperbolic discounting is a bias studied in behavioural economics. *Find citations explaining hyperbolic discounting*

A hyperbolically discounted score for an event after a delay D with discounting factor k is calculated by:

This score can be applied to a history of games played with different character pairs to weight that

history; this weighting can then be used to create a new PMF to draw character pair choices from, by allocating each pair a probability of being chosen according to its share of the weighted history of winning pairs. Where the previous method for selecting character pairs selected each pair with the frequency they appear in the record of winning pairs, this method weights the record of winning pairs; the sum weighted history of a character pair's wins as a proportion of the total weighted history for all pairs becomes the PMF for character pair selection. *Maybe there's need here for a diagram or some extra equations to make this clearer...?*

With this scoring mechanism, our model of learning incorporates a behavioural bias which some real-world players may exhibit. Player character choices can now be modelled through different behavioural variations applied to the naive model. The resulting behaviours from each model of learning can be compared to that exhibited by the naive model to discern which produces simulated RPG Lite play more similar to a given real-world player's; methods for doing so are explored in the remainder of this chapter.

6.5 Experimental Design

The models of learning described in section 6.4.3 can be applied to augment the naive model described in section 6.2 and introduce change in player behaviour over time. The remaining challenge is to demonstrate that augmenting the naive model can make it more realistic in practice. To that end, the remainder of this chapter details an experiment which compares synthetic datasets produced by the model to empirically sourced data, and shows that this synthetic data is more similar to its empirical counterpart than that produced by a naive model with no aspects applied. *Some reference to the research question is required here.*

This section presents classes used in the design and implementation of experiments, as they constitute the foundation of the execution and evaluation of synthetic datasets. This is followed by documentation of the strategy for generating synthetic datasets, and a description of the technique used to compare them to select optimal parameters for player simulation. This design is used in the experiment itself, described later in section 6.6.

6.5.1 Optimising the Model to Simulate a Real-World Player

I need to write up the strat for using k-fold validation. Should include the parameters selected, evaluation of “best” result, etc.

Discussion required here around the mitigation of randomness — it’s required because of RPGLite’s inherent RNG — and also explaining how we make use of a grid search in combination with k-fold validation to anneal toward well-fitted parameters of our models of learning for each individual player.

TODO make a diagram of this.

6.5.2 Model parameters

Many parameters are used to control simulated play of RPGLite. Their use in modelling learning or running simulations for data generation is explained in following subsections; however, parameters which control an experimental run are collected in a class, `ModelParameters`. Defining these values in a class helps to identify them in the codebase, and defines all variables which are used when optimising the model. These parameters are required because we run with different parameters for different folds. This subsec should be preceded by another on our strategy for optimising, so we can introduce k-fold validation and clean up the wording in this sec so it makes more sense.

The `ModelParameters` class is shown in fig. 6.4. The parameters it defines, briefly explained, are: IIRC Tim doesn’t love the description list formatting. Probably means I should avoid; consider other ways of laying this out that avoids a big wall of text

starting_confidence The initial value of the model’s confidence curve.

assumed_confidence_plateau A corresponding high value for the model’s confidence curve. As the curve’s growth rate is calculated to align with the number of games completed by the real-world player being simulated, and the curve asymptotically approaches 1, a high value is required which represents the expected confidence of a real-world player having played a significant number of games.

```

1 from dataclasses import dataclass
2
3 @dataclass
4 class ModelParameters:
5     c: float
6     curve_inflection_relative_to_numgames: float
7     prob_bored: float
8     boredom_enabled: bool
9     training_data: list
10    testing_data: list
11    assumed_confidence_plateau: float
12    starting_confidence: float
13    iteration_base: int
14    number_simulated_players: int
15    advice: list[tuple[str, str, str|callable]]
16    players: list[str]
17    args: list[any]
18    kwargs: dict[str: any]
19    boredom_period: int = 25
20    initial_exploration: int = 28
21
22    @property
23    def boredom_period(self) -> int:
24        '''
25        Number of games to play before checking player boredom.
26        Attempts to allow every player combo to play each other once on
27        average before checking again.
28        '''
29        return int(self.number_simulated_players**2)/2
30
31    def active_dataset(self, testing) -> list:
32        return self.testing_data if testing else self.training_data
33
34    def iterations(self, testing) -> int:
35        '''
36        The number of games to play when generating a synthetic dataset
37        '''
38        if self.boredom_enabled:
39            return self.iteration_base
40        return int(self.number_simulated_players**2 * len(self.
41            active_dataset(testing)) / 2)
42
43    def rgr(self, testing) -> float:
44        '''
45        RGR for the parameterised c value, number of games to play, and
46        start/end confidences.
47        '''
48        if not hasattr(self, '_rgr_cache'):
49            self._rgr_cache = dict()
50        if self._rgr_cache.get(testing) is None:
51            num_games_to_confidence = len(self.active_dataset(testing)) *
52            self.curve_inflection_relative_to_numgames
53            self._rgr_cache[testing] = \
54                curveutils.rgr_yielding_num_games_for_c(
55                    num_games_to_confidence,
56                    self.c,
57                    start=self.starting_confidence,
58                    limit=self.assumed_confidence_plateau)
59        return self._rgr_cache[testing]

```

Figure 6.4: A class defining the parameters of the model of learning RPGLite which vary when optimising for a given player. For layout & space reasons, some methods of minor significance are removed.

curve_inflection_relative_to_numgames The proportion of games played by the real-world player being simulated at which the player's confidence curve should reach the parameter `assumed_confidence_plateau`. This allows for control over the growth rate of the curve while linking the rate of growth to the number of games played by the real-world player being simulated;

C The curve parameter of the confidence model's underlying birch curve;

prob_bored The probability that a player with confidence above `assumed_confidence_plateau` becomes "bored" and stops playing RPGLite. These players are removed from the simulated playerbase and replaced with new players. This mechanism is discussed in further detail in section 6.5.5.

boredom_enabled Whether the boredom mechanism discussed in section 6.5.5 is to be enabled.

training_data The set of real-world games used as a training fold when optimising parameters for simulating a player.

testing_data The set of real-world games used as a testing fold when optimising parameters for simulating a player.

iteration_base The number of games to play when generating a synthetic dataset. If `boredom_enabled == False`, the number of games is instead calculated to ensure that all synthetic players play the same number of games as were completed by the player being simulated, (calculated from the training or testing fold as appropriate).

number_simulated_players The size of the simulated playerbase.

advice A list of tuples defining advice to weave when generating data. Pieces of advice are uniquely defined by a tuple containing the type of aspect to weave (such as "`prelude`" or "`error_handler`"), a string-represented regular expression defining the join point of the advice, and a `callable` to use as an aspect when weaving the advice. As `ModelParameters` instances are serialised to disk to persist experiment results and Python functions are not serialisable, this may also be a string value; strings are IDs of aspects, and are replaced with their corresponding aspect on deserialisation.

players The player being simulated. This parameter is a list of player usernames to support annealing to multiple real-world player behaviours if required.

```

1 class ModelParameters:
2
3     # <snipped additional attributes and methods for space>
4
5     def run_experiment(self, testing, correlation_metric):
6         real, synthetic = generate_charpair_distributions(\
7             rgr_control=self.rgr(testing=False),
8             iterations=self.iterations(testing=False),
9             games=self.training_data,
10            players=self.players,
11            birch_c=self.c,
12            sigmoid_initial_confidence=self.starting_confidence,
13            boredom_confidence=self.assumed_confidence_plateau,
14            num_synthetic_players=self.number_simulated_players,
15            boredom_period=self.boredom_period,
16            prob_bored=self.prob_bored,
17            advice=self.advice,
18            *self.args,
19            **self.kwargs)
20         return Result(self, real, synthetic, correlation_metric, testing)

```

Figure 6.5: The `run_experiment()` method on the `ModelParameters` class, which can be used to reproduce an experimental run from its parameters.

args Any additional arguments to pass when running a simulation.

kwargs Any additional keyword arguments to pass when running a simulation.

boredom_period The number of games completed by the simulated playerbase between checks of players' confidence and random selection for removal (default 25).

initial_exploration A value which controls the number of games individual players complete before their confidence & learning models control their character pair selection (default 28, to allow sufficient games to select every available pair exactly once).

These parameters are varied when optimising the model to represent the learning of individual players in order to find a combination of parameters which best represent a given player.⁸ A `ModelParameters` object defines everything required to run a simulation of `RPGLite` play and produce the resulting character pair distribution for analysis. Individual experimental runs⁹ can therefore be reproduced using the class. The convenience method `run_experiment()`, as shown in fig. 6.5, is provided to simplify reproducing experimental runs with `ModelParameters` objects.

⁸The strategy for this optimisation is described in section 6.5.1.

⁹Throughout this chapter, "experimental run" is used to refer to the process of producing a simulated dataset for comparison against training or testing folds taken from the dataset of completed `RPGLite` games. Many experimental runs with different parameters are used to optimise a model.


```

1 from dataclasses import dataclass
2 @dataclass
3 class Result:
4     params: ModelParameters
5     real_distribution: list[float]
6     simulated_distribution: list[float]
7     correlation_metric: Optional[Callable]
8     testing: bool
9
10    @property
11    def pval(self) -> float:
12        return self.correlation_metric(self.real_distribution, self.
            simulated_distribution).pvalue
13
14    @property
15    def statistic(self) -> float:
16        return self.correlation_metric(self.real_distribution, self.
            simulated_distribution).statistic
17
18    def within_acceptable_bounds(self, pval_threshold: float,
        statistic_threshold: float) -> bool:
19        return self.pval < pval_threshold and self.statistic >
            statistic_threshold

```

Figure 6.6: The `Result` class used to collect results of experimental runs and evaluate their success.

6.5.3 Representing the result of an experimental run

As referenced in fig. 6.5, a `Result` class is defined to collect the results of an experimental run and simplify analysis of the distribution of character pair selection produced. A definition of the `Result` class is given in fig. 6.6. This class is used in the analysis of the success of an experimental run when selecting ideal experimental parameters for the simulation of a real-world player.

A `Result` is not the result of an entire experiment, or the ideal parameters to simulate a given player. Instead, it represents the result of attempting to generate realistic data for a player, which is defined by a `ModelParameters` attribute. Several `Result` objects can be compared against desired probabilities of correlation and correlation statistics (the `pval` and `statistic` properties, respectively) to select an optimal set of parameters for the simulation of an `RPGLite` player.

6.5.4 Quantifying Similarity of Character Pair Selection

This subsection most likely needs to be condensed. The experiments presented in this chapter make comparisons of many synthetic datasets against a single empirically sourced dataset. The latter

contains all games completed by a player whose learning of RPGLite we aim to simulate, and is drawn from the published RPGLite dataset [KWM20]. Comparisons are required to determine which of the synthetic datasets most closely matches the empirical dataset. We are specifically interested in determining whether players played with preference toward similar character pairs. The distribution of their choices describes a bias toward some pairs, and away from others. The distribution of the naive model, being randomly chosen, should exhibit no bias; a simple null hypothesis is therefore that a simulated dataset is as similar to real-world play as a randomly generated one; there exists no bias comparable to that exhibited by the empirical dataset in this case. Our hypothesis in answering the first research question posed is that, with models of learning applied through aspect orientation, the naive model can exhibit biases toward character pairs similar to that of a real-world dataset. The simulated play would therefore exhibit a more “realistic” behaviour in character selection than the naive model as a result of the aspects applied.

We expect discrepancies in the specific character pairs selected in games (such as the order of character selection) because of the random nature of character selection. However, it is feasible to assess the similarity between the distribution of character pairs chosen by simulated and real-world players. The distribution of a player’s chosen character pairs describes the preferences of that player, as it allows their selections to be ranked. We can examine the similarities between a player’s character pair preferences and the preferences of simulated players by ranking and correlating their preferences to arrive at a measurement of similarity between the datasets; this is achieved through rank correlation.

Many rank correlation measurements exist. For the purpose of comparing lists of character pairs, we require a non-parametric measurement. Some methods for calculating correlation assume a distribution of data, or require data to be distributed in a consistent manner; as nothing is known about the distribution of character choice preferences (and this distribution could vary between players), a measurement which is independent of the underlying distribution is required. Two common options are Spearman’s ρ and Kendall’s τ . **Find citations for each correlation coefficient, and also for justifying choosing between these two specifically (if required).** These measurements assess correlation in different ways, which can affect measurements of character pair preference. Spearman’s ρ accounts for the magnitude of difference between ranks, and is calculated as $\rho = 1 - \frac{6\sum(d_i^2)}{n(n^2-1)}$, where d_i^2 is the difference in the rank of the i^{th} observation in each dataset and n is the total number of datapoints in a

dataset. Kendall's τ accounts only for their ordering. It can be calculated as $\tau = \frac{C-D}{C+D}$, where C is the count of concordant pairs, and D the count of discordant pairs.

To choose an appropriate measure, we relate them to potential patterns in our datasets. As we expect many character pairs to be infrequently picked (if at all) once players develop preferences, a slight increase in selecting uncommon character pairs within a simulation could incur a significant increase in rank. This would have a more significant impact on ρ than on τ , as the larger difference in rank is squared in the calculation. The simulated dataset might also produce a distribution where rankings between empirically sourced and synthetic datasets had small differences, but were similarly positioned throughout the dataset. This would have an outsized impact on τ , as relative positioning affects concordance, but the square of a small difference in rank remains relatively small, meaning that ρ would be less affected by this pattern. A choice of correlation metric is thus a choice of what correlation means: should correlation more closely reflect preference of character pairs relative to each other, or “forgive” small relative preference deviations but impose a greater cost on the scale of disagreement on a given rank?

We choose to prioritise the preservation of rank order (τ) over the minimisation of rank difference (ρ). Unfavoured character pairs' ranks might be affected greatly by a small difference in the number of games played, because unfavoured character pairs play few games by definition. A small difference in selection count is therefore more likely affect rank; however, the character pairs would still be unfavoured relative to a players' preferred pairs, which are expected to have a disproportionately high number of games, also by definition. Minimising rank difference is therefore sensitive to RPG Lite's random nature. Preserving the ordering of character pair selection is also important when considering favoured pairs, as differences might indicate a difference in patterns of play; however, small differences in the ranks of favoured pairs with ρ would have a less significant impact than when measuring rank ordering with τ : ρ accounts for the square of the difference in rank, which remains small for small differences.

For these reasons, we quantify similar character pair preferences by representing preference as an ordered list of string-represented character pairs, and measuring their correlation using τ . I'm not sure I've defined how I represent (or calculate) the ordered list of preferences yet. If not, that should live somewhere, I suppose.

6.5.5 Controlling State Space Exploration

The model of confidence used grows monotonically as described in section 6.3.2, which affects simulated players' exploration of possible character pairs. Models of learning use confidence as a probability that their alternative behaviour is to be encoded; therefore, when confidence is high, players are likely to select character pairs to play which were observed to be likely to win historically. As character pairs which are not played cannot win games, pairs which are not selected for play cannot be chosen by models of learning which rely on a players' historical observations, which could lead to inaccurate simulation. To counter this possibility, two mechanisms are introduced to control state space exploration: "initial exploration" and "boredom".

A little clumsy — reword this para? Is it sufficient? Initial exploration is a threshold number of games a simulated player must complete before models of learning make changes. Learning models' impact in player behaviour is gated on both a player's level of confidence and the number of games played. This allows players to observe enough win/loss data in their early games that later decisions directed by the learning model have sufficient historical observations to have explored the state space make "informed" decisions. The length of the initial exploration phase for any player is given as an integer number of games by the `initial_exploration` parameter when generating data, as mentioned in section 6.5.2; use of the parameter when generating data is discussed in section 6.5.6.

Boredom is a mechanism which solves a problem introduced by some combinations of model parameters: if the confidence model's growth rate and curve parameter are high, player confidence raises sharply, with the effect that simulated players are unlikely to explore the state space. A situation can arise where the simulation's metagame falls into a local optimum, where the entire playerbase collectively observe many historical wins for some character pairs because of a disproportionately high rate of random selection when players explore the space of possible character pairs. This can be overcome by introducing new players into the playerbase over time. New players are guaranteed to explore the state space due to the initial exploration phase, and will continue to do so while their confidence grows. The randomly selected character pairs played in this time means veteran players will play games which would not be selected otherwise. This reinforces existing preferences when novel character pairs are weaker than veteran players' preferences, and weakens existing preferences

At time of writing metagame is not defined, but this should be introduced in chapter 5; there's a corresponding revision note there.

if veteran players are exposed to novel character pairs which are likely to win games played against their preferred pairs. In disrupting a locally-optimal search of the space of character pairs, boredom mitigates potential issues with random play, while allowing simulated players to learn over time.

Boredom is implemented using several parameters. Players are eligible to be removed from the playerbase when their model of confidence reaches a threshold, defined by `assumed_confidence_plateau` as mentioned in section 6.5.2. Once this threshold is reached, each player is removed with probability `prob_bored`, allowing some confident players to remain in the playerbase for a long time. **TODO note that players are removed, and then replaced, so that the playerbase is of constant size but varying experience! Replacement happens immediately. Mention boredom period controlling how often we check that players are bored. TODO note that boredom can be disabled**

6.5.6 Producing Datasets

Our goal is to produce datasets of simulated player behaviour with aspects applied. This is achieved by weaving aspects to augment behaviour and repeatedly invoking the naive model's `play_game` function while managing the simulated playerbase.¹⁰ The experimental codebase contains a function which is responsible for these tasks, `generate_synthetic_data`, which is parameterised as follows:

- `rgr_control:float`, which is the RGR value provided by the `rgr(testing:bool)→float` method of a `ModelParameters` instance. This is passed to `generate_synthetic_data` directly to maintain ignorance within the function as to whether it is training or testing against a dataset; the caller is aware instead, and passes the correct RGR value;
- `environment:dict()`, the environment simulated players interact with, which is used to accumulate information shared by all players such as a history of games or special values used to parameterise aspect behaviour;
- `params:ModelParameters`, the `ModelParameters` instance which provides parameters detailed in section 6.5.2 and is used to parameterise aspect behaviour and the model of boredom used to manage players;

¹⁰For example, generating synthetic datasets involves

I end list items here with semicolons. I should be consistent through the thesis, and decide whether semicolons are appropriate.

→ `print_progress:bool=False`, a parameter controlling the printing of debug loglines.

This is sufficient to generate a synthetic dataset, as all values which parameterise the behaviour of simulated players are encapsulated in the `ModelParameters` class.

6.5.7 Answering the first Research Question

In designing advice which amends a model with new behaviour, the first research question was effectively answered:

“Can changes to a model be represented as advice using aspect orientation?”

The advice described in this chapter demonstrates that behavioural variations such as learning can be separated into aspects and applied to a model to alter its behaviour without requiring its original implementation to be modified. We can therefore answer this research question: *yes*, modifications to a model can be represented using advice.

A note here that we’re using AOP, but don’t actually target cross-cutting concerns in our models of learning, which is interesting.

Other research questions motivated in the literature review in chapter 2 are concerned with whether representing changes to models using aspect orientation can accurately reflect their subject, and whether the modules formed by a library of aspects can be reused. These are investigated in section 6.6 and chapter 7 respectively.

6.6 Experimental Results

To be honest, this might be the best explanation of the experimental design I’ve produced so far. Maybe I should be re-using this in the experimental design section?!

The second research question proposed in section 2.4.1 is:

“ Can a model be made more realistic by applying aspect-oriented improvements?”

To investigate this, the naive model of RPGLite play described in section 6.1.2 is augmented using the aspect-oriented models of learning outlined in section 6.4. The datasets produced by executing both the unaltered naive model and the naive model with aspects applied are compared against the real-world dataset described in chapter 5. Comparisons are made by quantifying the similarity of the character pair preferences found in each dataset as defined in section 6.5.4. Whichever synthetic dataset is most similar to the empirically sourced one can be said to be the most realistic. This experiment has the null hypothesis that introducing aspects has no impact of model realism. If no change in model behaviour can be measured when learning models are applied, then it would be possible to *represent* model changes as advice as demonstrated in section 6.4 but not possible to *use* those changes in their aspect-oriented form.

This section evaluates the results of each model of learning individually. It starts by defining the comparison between generated and real-world datasets in more detail. With this background, an evaluation is presented of the aspects implementing character pair selection from the known (real-world) dataset, followed by an evaluation of the aspects implementing confidence-based models of learning and hyperbolic discounting.

6.6.1 Identifying Statistically Significant Results in Datasets

A discussion of how datasets are evaluated. We produce *many* datasets which have annealed to the highest stat and P val we could find, but they're split across many folds. How do we determine whether a set of parameters works "enough"? We find the highest correlation val and lowest pval which work across a majority of folds, then test against the complete dataset to be absolutely certain that correlation doesn't drop. This gives us the best parameters for a player: the parameters were reliably annealed to across many folds to give statistically significant correlation, retained their correlation against a testing set, and further retained their correlation against the entire dataset.

Is there a pattern in players getting reliable correlation and having played many games? Maybe we just don't have enough games to do this for most players. Check this; could be worth a note toward

the end, in a discussion or similar.

This is an explanation of the analysis script, not the process of annealing; we should already have that in the experimental design section (currently 6.5, at time of writing).

6.6.2 Selecting Character Pairs from a Known Distribution

The second research question presented in section 2.4.1 is:

“ Can a model be made more realistic by applying aspect-oriented improvements?”

This question can be answered directly by weaving advice which modifies player behaviour so that character pairs are selected from a distribution calculated from existing player data. If the dataset produced with this advice woven correlates with the pre-calculated distribution of character pair preferences, then the simulated behaviour must have been affected by the aspects in the manner the advice modelled, and so made the model more realistic. If the dataset correlates with that produced by the naive model, it must have had no effect. And, if it correlates with neither, then the weaving of advice modelling a specific character pair selection did impact player behaviour, but did not produce expected results.¹¹

Datasets representing every real-world player who completed more than 100 games in RPGLite’s first season were generated with the advice described in section 6.4 woven. There were 21 such players. The correlation statistic (and corresponding p-values) of these datasets compared to their players’ real-world datasets and the naive model’s dataset is presented in fig. 6.7.

<i>Simulated Player</i>	<i>P-Value</i>	<i>Correlation Statistic</i>
Data1	data2	data3
Data1	data2	data3

Figure 6.7: Correlation statistics generated by weaving aspects into the naive model which select character pairs with the same distribution as a specific real-world player

TODO: add notes explaining what an acceptable pval and correlation stat are. This will be useful to refer back to in the next subsection, where not all players are simulated by each model.

¹¹In this scenario, there would likely be an error in implementation as the aspects demonstrably made a difference to behaviour, but not what was intended. Rather than confirming the hypothesis that advice can be used to augment models of behaviour or the null hypothesis that weaving such advice has no effect, it would indicate a flawed experiment.

Explain the results & relate them to the final RQ.

6.6.3 Selecting Character Pairs using Confidence-Based Learning

I suspect I'll move this to the discussion of the next experiment, so that the aspects using the known distribution are presented as an investigation of the first research question, and the aspects using the confidence-based model of learning are presented as an investigation of the second.

Selecting character pairs based on a known distribution demonstrates that advice can alter the behaviour of actors in a model, improving the model's accuracy.

I'm increasingly realising that the aspect drawing from a known distribution is both more accurate and more portable than our models of learning: we can apply it to S2 and just draw from season-specific data to get the correct results. It's actually way better than our models of learning. However, the models of learning let us investigate specific players' behaviour, so they let us use the naive model to investigate the behaviour of real-world players, which is still quite valuable, but might want a different approach to the write-up. I'm pretty lost with this chapter — need Tim's input, I think.

6.6.4 Learning with Hyperbolic Decay

6.7 Discussion

A closing discussion on what we found, and how the research question was answered.

Also, note the weaknesses of the approach.

- More games could have lead to a more accurate result
- It'd be interesting to use the models to "learn" other things too, such as move selection / mistakes made
- Would like to reproduce across multiple different systems, not just RPLite. Maybe developers like we did in the CAISE paper, maybe aspects that aren't learning like degraded modes. Could

be interesting to model a complex adaptive system with aspects representing rule changes. Could be interesting to demonstrate other uses of aspectually augmented systems, like introducing new rules to a cellular automaton.

Chapter 7

A chapter title here for the experiment moving aspects to new systems, or systems with some changes

This is a relatively short chapter; a lot of the building blocks for it exist in the previous chapter, so there's less ground to cover. If it ends up quite lop-sided, I'd chop the earlier chapter in two rather than artificially making this chapter beefier; I think it'd flow better.

7.1 (Reword) Experimental motivations / motivation of research question

This section should describe why it's interesting to move an aspect trained on some actors' behaviours to a new system, and discuss what investigating this can teach us that we *don't* already know from the previous experimental chapter. We've got aspects which represent behavioural variance. If the system changes, can we expect that these aspects still apply to the new system? Is the representation of behavioural variance in this model separable from the system the behaviour occurs within?

7.1.1 Coupling of Model and Behaviour

Explaining the issues of behaviour coupled to models. Highly related to some of the reviewed literature — I forget exactly what — which discussed whether aspects which were designed to be applied to one system could feasibly be transferred to other systems, or whether they're inherently aware of the system they're originally designed for. The core concept is that to be making changes to some underlying codebase, you probably have to know what that codebase is, in most cases at least.

7.1.2 (Reword) Gaps/opportunities left by previous experiment

In the previous experiment we demonstrated that behavioural variance can be plausibly realistic. Are those variations separable from their underlying model, or — when trained i.e. made realistic — do they suffer from the coupling discussed in the previous subsection?

7.1.3 Research Question

Whatever the specific research question's wording for this was. Something about decoupling realistic aspects from a given model maybe? It should be in an earlier chapter somewhere.

7.2 Experimental Design

We took old player data and trained aspects on them, and in the previous experiment we found they were statistically significantly accurate. We ran a second season of RPGLite with slightly different parameters on player data. We modelled how players learned and variations on their learning patterns, so in theory, we should be able to apply the new learning patterns to the other system too. Do we have to re-train the aspects? How portable are they? This section lays out the design of this experiment.

7.2.1 Changes to RPGLite

What changes did we make to our system?

7.2.2 Applying Behavioural Variations to New System

Layout of new experiment, how it'll work, what's measured, why it should answer the RQ.

7.3 Applying Aspects from a Control System to a New System

Our implementation & results from the experiment described above. Discusses how the above experiment was realised, lays out the results we found, and relates those results to the research question we started with.

7.3.1 Implementation

Implementation of the experiment

7.3.2 Results

Presentation of the results of the experiment, analysis, some discussion (more in next section too)

7.4 Discussion

Some notes discussing the outcome of this research question (that trained aspects representing behavioural variance aren't separable from the system they're trained around)

Chapter 8

Future Work

The focus of this thesis is to develop a state-of-the-art aspect-oriented framework, to produce a suitable experimental environment to demonstrate its effectiveness, and to use that environment to investigate whether aspect orientation can be used to augment models with behavioural variance. *is suitable for simulation purposes? It's about showing that we *can* use aspect orientation appropriately in simulation environments, and that aspect orientation can also lead us to realistic and nuanced simulations, too. Go back through the pdf and lit review chapters to make this argument properly.*

. We have found that aspect orientation can be used to create realistic and nuanced simulations, have successfully produced a well-constrained environment to simulate for testing purposes, and have produced a novel aspect orientation framework which demonstrates novel and powerful weaving concepts; lots of opportunities for research outwith this thesis' scope present themselves.

This chapter describes some possibilities for the presented research to be extended in the future.

Add a RQ at the end of each research opportunity; this can and should be a contribution of the thesis, and the RQs would help to sell it as such.

8.1 Aspect-Oriented Metaprogramming in real-world Software Engineering

The combination of metaprogramming and aspect-orientation introduces new possibilities in aspect-orientation. In traditional aspect-oriented work, aspects treat their targets as black boxes; this leads to limitations which aspect-oriented metaprogramming can address.

Traditional aspects cannot intersperse their behavioural modifications with the work being done by their target, as they apply their logic before or after the target's execution. The "textbook" use case for aspect-orientation is logging: aspects can separate logging from the business logic they are applied to. However, a programmer in mainstream programming paradigms may wish to insert logging behaviour *within* their business logic rather than *around* it. Aspect-oriented metaprogramming enables this as the target can have logging logic woven within business logic, which remains decoupled within the codebase. To achieve this end goal without "within"-style aspect application would require a refactoring of business logic to create join points which traditional aspects could apply against; this breaks the "obliviousness" property of the aspect orientated philosophy discussed in section 2.0.1.

8.2 Augmentation of pre-existing models

Models which predict the future state of some system may be accurate in general, but cannot account for unforeseeable events such as financial collapses, pandemics, or unpredictable weather events. Pre-existing models cannot account for these one-off shifts in system state due to their random nature. For example, models of world health over time could not account for the Covid19 pandemic, and models of local or global economies could not incorporate real-world data from the recession resulting from the pandemic. The 2008 financial crisis also impacted local and global economies, but could not be predicted before the event. The World3 model is an example of one which has provided accurate high-level predictions of the state of various global systems over many decades [Find citations for World3 model](#) , but predictions for recent years fail to account for these unforeseeable events.

If the modification is represented as a special case within the model, a research software engineer must introduce their corrective code alongside the model's core logic, thereby tangling¹ the two. It

¹For an overview on the tangling of cross-cutting concerns, see section 2.0.1.

therefore has the properties of a cross-cutting concern [Kic+97; FFN00], and can be factored into an aspect as such.

Research investigating the use of aspects to correct for one-off events could aim to show that an aspect can “nudge” a system’s state in-line with real-world data when exceptional circumstances affect a modelled system but cannot be elegantly represented within the model’s logic. A research question for this future work is therefore:

“ Can aspect orientation be used to introduce special cases to real-world systems to correct a model’s predictions in when a system under research is altered by a freak event?”

lorum ipsum

“Freak event”
might be too
informal here
but I’m not
sure what
to replace it
with.

8.3 Aspect Orientation's utility for Research Software Engineers

A corollary of the research opportunities for aspect orientation’s use in software engineering discussed in section 8.1 is that there is an opportunity for research software engineers to benefit from the adoption of aspect oriented programming. However, while aspect orientation’s use in industrial software engineering has drawn criticism [Ste06; Prz10; CSS04]² its use within research codebases is a special case where it may be more suitable. Related suggestions for the use of aspect orientation in research codebases were proposed by Gulyás and Kozsik (as discussed in section 2.2.1). However, the benefits proposed in earlier work concerned the design of the software itself. Aside from aspect-orientation’s utility in software design in a research context, there are potential benefits for the *practice* of developing these codebases which may be of professional interest to research software engineers.

The resource constraints and stringent requirements for accuracy in research codebases present challenges which aspect-orientation could assist with. Compensating for a special case in a pre-existing model would require maintenance of the codebase, which takes time and could inadvertently alter its behaviour. Time is a scarce resource in research environments, and undesired changes to a model’s behaviour can invalidate research results. An alternative to adjusting the models directly is to construct aspects which represent large and unpredictable events in real-world systems such as pandemics,

²A review of aspect orientation’s critique is presented in section 2.0.2.

economic crises, war and famine. These can be modelled on real-world data for accuracy, which can produce realistic simulations as this thesis demonstrates. This use of aspect orientation in simulation and modelling can be investigated by creating a proof-of-concept of the approach as applied to pre-existing models.³ Studies can also be conducted to investigate whether an aspectually-augmented model is quicker to construct and easier to maintain in future than a codebase with “patches” written into its original logic.

Researchers investigating this technique’s application to existing models could also investigate the difficulties of augmenting a model created with no intention of weaving advice in the future. The construction of advice requires appropriate join points to be specified, and codebases which are structured in a way which doesn’t yield convenient join points might be more complex to augment aspectually. These cases raise another use-case of PyDySoFu’s “within”-style weaving through runtime metaprogramming: where other aspect orientation frameworks force aspects to treat the targets they are invoked on as black-boxes, PyDySoFu can make modifications within them. PyDySoFu’s new kinds of aspects may therefore be more useful than those in other frameworks for researchers responsible for maintaining codebases written with no consideration for join-point specification, and would therefore be more universally applicable; this possibility requires future investigation.

If the research described successfully shows that aspectually-augmented simulations are easier and quicker for an RSE to maintain and deliver than direct maintenance of the model’s codebase, then augmentation of existing models to improve their accuracy can follow in the community.

If I’ve got time, I’d love to push my codebase augmenting a World3 model to achieve just this with predictive implications for the 2008 financial crisis and the covid pandemic.

8.4 Hypothesising Possible System Dynamics via Aspects

Unpredictable events can cause discrepancies between simulated system state at a given time and the real-world system it models, as discussed in ???. While this technique presents promising research opportunities, researchers face other kind of model uncertainty which aspect orientation could also counter.

In some scenarios, it is difficult or impossible to make predictions about a system’s future states because its dynamics are actively being researched. Standard scientific practice is to create a model —

³This approach is similar to PyDySoFu’s initial proof-of-concept study [WS18a] as discussed in chapter 3.

mathematical, computational, or otherwise — to create synthetic datasets which indicate accuracy if their predictions align with what is empirically observed. **Cite popper, any point including studies that do this too? It's fairly table stakes as a concept...** Some aspects of the system under study may be well-understood.

Rather than creating models of an existing system which encode its hypothesised behaviour, a naive model can be created which operates as the scientific consensus understands it. Hypothesised behaviour takes the form of aspects altering behaviour in any manner the hypothesis requires. Within-style aspects allow arbitrary modifications, to simulated behaviour, increasing the flexibility of the technique. Data produced by each model can be compared to empirically sourced data, and their similarity quantified, as demonstrated in chapter 6 and chapter 7. Our null hypothesis is that the naive model's similarity to empirically sourced data is greater than that of the aspectually augmented model; our experiment's hypothesis is that the behavioural change applied as aspects is more representative of the system under study than that of the community consensus.

This technique has a satisfying property: the hypothesis in a given experiment is clearly described by its aspectual representation, and if an experiment is successful — meaning its hypothesised behaviour accurately represents the real-world system — then the model adopted in future research as that of the community consensus can be the aspectually-augmented one produced by the original study. In this way, the scientific process is directly represented in the structure of the codebase, and the community's progression to increasingly accurate models of a system is represented by the progressive adoption of “patches” to an original theory.

Maybe too informal, but I don't know, maybe live a little...?

Hypotheses can also be created compositionally in this model. Researchers might develop a series of potential system properties or behaviours, but are unable to investigate all reasonable combinations in a timely manner. However, sets of aspects representing each can be composed to produce, for N hypothesised behaviours, 2^N potential behaviours which can be compared to empirical datasets to identify which combination of possible behaviours most closely resembles that of the real-world system under study. The technique is similar to modern pharmaceutical development, where a series of algorithmically identified compounds can be produced and tested to identify which are most suitable to treat a medical issue **find a citation for modern algorithmic / robotic drug development**, but can be applied to any models where aspectual augmentation is appropriate.

This sub-sec is already quite long, and including this

This technique for developing experimental model codebases has another desirable property: it unifies seemingly incompatible philosophies of the scientific process. Kuhn [Kuh12] explains the scientific process as inherently social: it starts with a paradigm which is accepted as broadly true, and accumulates an increasing number of exceptions until the paradigm itself is deemed unfit, and a new basis for a field's research is adopted. Popper [Pop13] explains the scientific process as an approximation towards truth, with incremental progress made with each result achieved by a research community. The proposed technique for developing experimental model codebases demonstrates features of both. Consider the original model a paradigm initially selected by community consensus, and the application of aspects Popper's incremental movements toward truth (as successful experiments are conducted) or Kuhn's exceptions to the agreed model (as experiments identify weaknesses in the original model). In this case, each successive new model adopted by a community is adopted in a Popplarian manner: improvements are objectively measured, incremental, and would trend towards truth as a model's behaviour fits empirical observations increasingly closely. However: over a sufficient period of time, the incremental patching of an original model would produce an accepted community model which contains a relatively large amount of discovery and complexity encoded in aspects, as compared to the original model they are applied to. One would expect the research community to rewrite the base model to simplify future aspect application and to more elegantly encode recent research findings; effectively discarding the original paradigm in favour of a new one. This process is Kuhn's "paradigm shift", where paradigms are dropped once a generation of researchers determine that an originally accepted theory on a topic is unfit for purpose as evidenced by mounting exceptions in the literature; a new paradigm is to be accepted by the community, as a new base model would have to be written and adopted.

The relevant philosophy of science is more nuanced than its brief explanation here, and the suitability of the approach for the development of research codebases is to be investigated; the work involved is outwith the scope of this thesis, but is suggested as future work. A basis for the incremental improvement of models via aspects is effectively demonstrated in chapter 6 and chapter 7, but the feasibility of the approach as a basis of a community's scientific process and relation to philosophy of science warrants further investigation.

Finally: hypothesised properties of a system might interact, meaning their discovery could involve

the fitting of multiple models of possible behaviours to discover the properties of a real-world system. The use of aspects to encode — and discover the presence of — hypothesised behaviour is discussed in section 8.9.3 as the RPGLite dataset presents convenient opportunities for doing so. The optimisation of multiple aspectual models of behaviour is discussed in section 8.8.

8.5 Standards for Aspect Orientation in Research Codebases

The possibility of a research community sharing models containing aspectual augmentation and possibly developing additional aspects to augment a model further involves a considerable amount of model logic written within as advice. The community developing these aspects have the responsibilities of maintaining a codebase, but the added complexity that research software engineering introduces. These codebases may be used for many years, and may be iterated in in a series of future experiments. The legibility of these codebases and their long-term maintenance are areas of criticism in the software engineering community [Ste06; Prz10; CSS04]. The research community must therefore mitigate these weaknesses of the aspect-oriented paradigm when adopting the techniques discussed in this chapter for simulation and modelling.

To address the concern of the visibility of advice being woven, researchers may already take advantage of :

- ① Improvements to tooling produced by the aspect orientation research community, including IDE integration [CCK03] and runtime inspection [MR02], should make clear to engineers what advice is being woven in a codebase and assist with debugging aspect-oriented programs respectively.
- ② Aspect-orientation frameworks specifically designed to clarify to an engineer the aspects being woven should allow for less friction on the part of a maintainer who inherits a codebase and must reason about its behaviour. This is particularly important if the maintainer aims to weave more aspects into the codebase, and so must understand its existing behaviour before augmenting it further. Adopting weaving patterns such as import hook weaving — described in chapter 4 — should make a program clearer to a developer regardless of the tooling they have access to.

I've made this an enumerated list; is it better as a paragraph?

The impact of framework design on a codebase's maintenance should assist a developer even in the

absence of tooling, but the success of import hook weaving in this regard is untested. An appropriate research question which arises is therefore:

“ Do aspect orientation frameworks with weaving techniques designed to simplify a developer’s understanding of a program affect a codebase’s long-term maintainability?”

8.6 Standard Aspect-Oriented Model Features

Researchers who build aspect-oriented models and extend others’ aspect-oriented codebases must be able to collaborate at least as easily as they currently do in a culture without aspect orientation. One way aspect orientation might improve researchers’ ease of collaboration is with standardised libraries for aspect construction. Similar libraries were developed when developing a case study for PyDySoFu’s viability [WS18a; SW16], but development was left incomplete as discussed in section 3.3.1.

The original library with this aim, Fuzzi-Moss, was originally designed to provide standardised aspects to represent behavioural variance in socio-technical systems. The aspects developed in Fuzzi-Moss have no notion of being fitted to real-world data. Instead, they are simple models of behavioural variances such as distraction, which are parameterised to allow users of the library to use these simple models in whatever manner is appropriate for their use-case. A broader collection of these behavioural variations could simplify the use of aspect-oriented behavioural variation in the research community writ large, by removing researchers’ burden to develop these themselves. Early construction of a library with Fuzzi-Moss’ goals would also support researchers in sharing models or extending others’, as they would be familiar with a common set of tools. The development of such a library and the production of case studies demonstrating its effectiveness would answer the research question:

“ Can researchers using aspect-oriented behavioural variance share a common set of tools to support and simplify the use of the technique in their codebases?”

Other tooling to support researchers in the development of aspectually augmented simulations and models could also be developed. For example, a library of fuzzers which make changes to an abstract syntax tree could be constructed. Such a library would not model specific behaviours, but would

allow researchers to build models performing within-style aspect weaving without writing code which contained no metaprogramming logic. Instead, this logic would be encapsulated in utility functions provided by the proposed library. This would also reduce the work required of researchers looking to use the techniques demonstrated in this thesis. The library could also support the development of a Fuzzi-Moss-like set of socio-technical behavioural variances. However, such a library does not currently exist. A summary the contribution to the research community which the proposed library would make is:

“ The development of a library of metaprogramming operations which simplify the construction of within-style aspects, supporting its research use and demonstrated in case studies.”

The proposed library need not be a separate codebase to the aspect orientation framework it is designed alongside. PyDySoFu could contain this collection of metaprogramming operations; however, there is no obvious requirement for the library to be contained within PyDySoFu, and separating the proposed library from the framework it is built to interact with removes any dependency on a particular aspect orientation framework; future researchers could develop alternative frameworks with within-style aspects which the proposed library could also be compatible with. We therefore suggest that the development of this library is a separate project from the development of PyDySoFu itself.

“obvious” is a little too informal, not sure what to replace it with here.

This might be a little rambly — consider omitting, rewording, or factoring into a footnote.

8.7 Testing Frameworks to detect Model Incorrectness

Many of the uses of aspect orientation in simulation & modelling research discussed in this chapter affect the construction and maintenance of a codebase, aspect orientation also has potential in the instrumentation of an experiment, as discussed in ??.⁴ Experimental instrumentation was demonstrated for data collection in this thesis’ experiments, as discussed in chapter 6 and chapter 7. **We can get more precise than chapter references; find the sections / subsections.** Instrumentation also has potential in other aspects of research software engineering; one possibility to be investigated is the use of aspects to alert researchers to impossible states visited by a model during runtime.

Many models construct a version of a real-world system which cannot achieve certain states. For

⁴In particular, see relevant notes on “The use of aspect-oriented programming in scientific simulations” [GK99].

example, a model of collisions might never be expected to see an increase in the total energy present across all modelled entities; simulations of socio-technical systems might expect bounds on the hours worked by a simulated workforce, or the amount of output the workforce creates; researchers might study a system to observe a n emergent property which is bounded by laws around its growth (for example, a linear increase in some property over time as opposed to exponential growth). It is important that any model of a real-world system accurately reflects the system's physical limits and underlying

mechanics. Models may fail to reflect these mechanics for many reasons, including :

Are there any other reasons as to why models might be incorrect (and instrumentation could identify this by boundary checking)?

- The model may be developed with bugs which are not detected, i.e. it is conceptually incorrect;
- The model may be verified as correct but later maintenance introduces undetected errors, i.e. it contains bugs;
- The model may be correctly written, but given unexpected inputs.

Software engineering practices such as unit testing are common practice and may catch these bugs through identifying incorrect behaviour in components. Thorough test suites may also include integration tests which observe the behaviour of many model components working together. However, there are limitations of both testing techniques:

- Test suites as described may fail to identify non-deterministic errors in a model, which arise only in a small number of cases;
- Traditional test suites also may fail to identify emergent state errors, where individual components operate correctly but their repeated integration results in an incorrect emergent property of the system ⁵;
- Test suites usually fix values such as random seeds or input data, which may hide non-deterministic state errors or may not exhibit the properties of real-world data (and so exhibit different behaviour under test than when an experiment is conducted).

We propose the use of aspects to instrument models to assert a system state within expected bounds, both as components of a test suite and as assurance that a model was correct at any time it

⁵Emergent properties are often the state under study, as in PyDySoFu's original case study [WS18a]

is used as part of an experiment. All of these limitations can be alleviated by a test suite which is able to assert the correctness of model state at different points in a program's execution. Models which non-deterministically reach erroneous states are primarily an issue during the execution of a model; aspectual state assertions need not alter a model's program flow, so are appropriate for use during its experimental use should any unexpected state arise. These states may also be deterministically caused by experimental input data which is not reflected by the inputs in a test suite; weaving these aspects during experimental use also addresses this limitation. Emergent properties of a system may also be tracked over time, and aspects may measure emergent properties to ensure they are within appropriate bounds at all times in an experiment and have correct properties of their own (for example, these aspects could ensure correct growth curves for the property in question).

Aspects can also be used as a regression test to detect the recurrence of bugs after they have been identified and fixed. If a simulated system enters an incorrect state during development, and the underlying bug is fixed, an aspect observing the system's erroneous property can be constructed which alerts developers if the bug is re-introduced.

Aspects instrumenting a model throughout its use in an experiment would also serve to guarantee peer reviewers of a study that the model was correct regarding the properties observed by the aspects. A suite of aspects which observe bounds on and characteristics of system properties is, functionally speaking, a declarative set of known-good behaviours of a model. Peer reviewers who sought to check that a model was correct should be able to extend this list of properties and observe any assertions during reproduction of a study's findings.

Within-style aspects are particularly appropriate for this task, as the measurement desired of a program might exist inside an existing function or block of code. It is conceivable that observing system state before and after a function executes would not be sufficiently granular to assert correct system state in some cases. Aspectual instrumentation of research codebases is therefore particularly useful in light of new weaving techniques in PyDySoFu, traditional aspect orientation frameworks would be sufficient in instrumenting at least some models for observation.

To demonstrate the efficacy of this approach, a library of tools to observe boundaries on system states could be developed, similarly to the library of tools suggested in section 8.6. Case studies

augmenting research codebases using either this observational tooling or ordinary aspects can then be constructed. These should include new codebases, to demonstrate the technique’s utility during development, but should also include the instrumentation of old models and the detection of any irregularities in existing, published work. Such a project would answer the research question:

“ Can aspect orientation be used to instrument a model and assert correctness by observing that the model’s properties remain within expected bounds?”

8.8 Optimisation of Multiple Aspects

Experiments in chapter 6 and chapter 7 made use of models which were fitted to real-world data. However, they only fitted parameters of learning models, and only some parameters were fitted. This was partly due to the computational cost involved in fitting these parameters and time constraints on

this research.

Optimisation is an independent research field with a broad range of techniques outwith the scope of this thesis to explore; however, a study of optimisation algorithms and technologies as applied to the fitting of aspects to real-world system models is a topic for future research. Models of real-world systems can be computationally expensive to complete, meaning that measuring the accuracy of a model with a variety of parameters could be computationally expensive. A survey of optimisation techniques suitable for aspectually augmented simulations and models would answer the research question:

“ Which optimisation techniques are best suited to fitting parameters of aspects woven into models of real-world systems, with regards the accuracy of the aspectually augmented model produced?”

8.9 Future Work pertaining to RPGLite

RPGLite’s dataset was analysed for the purposes it was collected for in this thesis: to aid in the realistic simulation of a well-controlled socio-technical system. However, some pieces of analysis have

Is this point worthwhile? I cited it in IIRC an experimental chapter, but I’m unsure it’s worth including. For future Tom to revisit.

not been completed; this section discusses future work enabled by the dataset of real-world play which this thesis contributes.

8.9.1 Causes of Game Abandonment

The games included in the dataset produced by RPGLite finish in varying states: some are incomplete (as some games were ongoing when the snapshot of the game's database was taken), and others are finished. A third category of games are ones where a player opted to abandon a game they were participating in. This feature of RPGLite was developed to allow players who were playing inactive opponents to free a game slot. It also allowed players who felt they had no hope of winning, or were bored with a particular game to forfeit.

As many datapoints are included in the published dataset, it is possible to observe how active players were (by measuring the frequency with which they would open RPGLite), and engagement with the app might be quantified by players' exploration of (or frequent use of) leaderboards and customisation features. Correlations might exist between players' activity, engagement, or experience with their propensity to abandon a game in different states. For example, one hypothesis could be that particularly active players frequently forfeited games against inactive players out of boredom, or that players who were unlikely to win a game (or were inexperienced players of RPGLite paired against competent players) might be more likely to forfeit. These results might be of interest to the broader game design community. Research investigating reasons for abandonment would answer the research question:

“ Can players who are likely to forfeit a game be identified through patterns of RPGLite play?”

8.9.2 Patterns of Play within Cliques of Players

RPGLite's application implemented matchmaking systems which allowed players to advertise themselves as open to new games, challenge players discovered on a leaderboard, or rematch against players they had previously played against through a game history. However, two factors might have

impacted how players chose their opponents:

- ① Recruiting methods means people might have known each other
- ② Players who find others they are well-matched against might repeatedly re-match rather than continuing to discover players they might have less satisfying games with

These factors would imply the formation of cliques of players: small groups who often created games against each other much more frequently than they did with the game's broader community. If so, this could have implications for uses of the RPGLite play dataset. For example, in the experiments discussed in chapter 6 and chapter 7 no structure is imposed on the matching of simulated players, so that players would be uniformly exposed to the game's community. As these experiments simulate learning from the outcomes of games, exposure to the entire community would expose players to learnings from all games played; cliques of players have a limited set of games to learn from, and so could affect what is learned by that clique. The broader community would also be insulated from what was learned within the clique, meaning that a loss of information affects all players.

Future research making use of the RPGLite dataset should be aware of any underlying patterns which could affect the dataset's emergent properties, such as learning styles. Clique formation is an example of such a pattern; others might yet present themselves. Future research making use of the dataset of RPGLite play may look to address research questions such as:

“Are patterns such as cliques of players present in the RPGLite dataset? If so, what patterns exist, and what is likely to have driven the formation of those patterns?”

8.9.3 Aspect Oriented Discovery of Dataset Properties

Relatedly, the dataset presents opportunities for research in aspectually augmented simulation & modelling related to the future work described in section 8.4. In section 8.4, the use of aspect orientation to encode hypothesised behaviour was discussed. However, in some scenarios, the expected behaviour may be unknown: it is currently unknown whether cliques were formed by RPGLite players; the behavioural impact of player interaction with ancillary game features such as leaderboards or

Consider refactoring this into a subsection of that sec.

profile customisation is unknown; players' early aptitude for RPGLite may cause them to lose interest relatively quickly due to a lack of challenge or spur them to climb leaderboards or continue to win games; players more likely to challenge others with more experience might learn and be more likely to continue playing RPGLite, or might quickly lose interest due to repeated loss.

These behaviours can be hypothesised, but many might have an impact on each other: players might find experienced peers through the leaderboard, so their progression in or engagement with the game could be due to their propensity to explore (by using the leaderboard) or their exposure to more advanced play (through games with experienced peers). This can be investigated through aspects fitted to real-world data, used to discover parameters for multiple hypothesised concerns in the player data by fitting them simultaneously.⁶ The relative importance of each concern could be discovered by inferring parameters controlling the strength of each concern which were fitted to real-world data. This would allow the existence and relative impact of real-world behaviours to be discovered.

This proposal is an extension of the research discussed in section 8.4, but the RPGLite dataset is a suitable testing ground for the technique which already exists. Larger-scale data collection described later in this chapter (see) could provide opportunities to test this more rigorously, as the discovery of nuanced emergent properties of RPGLite play might require a considerably larger dataset than this thesis had the scope to contribute.

Research into this technique would contribute answers to the research question:

“ Can potentially interacting emergent properties in a real-world system be identified through the fitting of aspectual models representing hypothesised properties to empirical datasets which may exhibit them?”

8.9.4 Larger-scale Data Collection

RPGLite's playerbase was recruited informally. Only two seasons of gameplay were run, and the game was not heavily advertised to prospective players, particularly once it was clear that sufficient data would be collected for the research that required it. As a result, there is scope for a larger and longer-term data collection effort to be made. A re-release of the application in major mobile app stores

⁶Research into relevant techniques is proposed in section 8.8.

with a concerted effort to release new seasons of the game and maintain player interest for an extended period of time — perhaps with additional features, such as in-game chat, favourites lists of previous opponents, or match replay and analysis (with suggestions for improved play backed by the formal methods inherent in RPGLite’s design) — would enable a richer analysis, and broader utility to the games research community.

RPGLite’s dataset contains information about players’ interactions with the application itself. As the game made available some features typical of modern games such as leaderboards, matchmaking, achievements, and graphical customisation, player data containing interaction data on features which are used commercially could be used to shed light on the more effective aspects of modern game design for engagement purposes. There are many opportunities available for the game design research community to investigate.⁷ Publications in the field from co-creators of RPGLite reflect further on the design and future improvements of the game [GameOn2020; KM20; Kav21; Kav+19; KM21].

A large-scale user study of RPGLite could also be used in pursuit of other research questions suggested in this chapter. Players could be classified by their likelihood of abandoning games using the patterns identified in section 8.9.1 to verify the patterns empirically. A larger dataset might be useful in testing the discovery of system behaviours using hypothesised behaviours encoded in aspects. This could be used for both the aspectual encoding of an experimental setup, as discussed in section 8.4, the discovery of features of player behaviour suggested in section 8.9.3, the discovery of dataset properties such as player cliques including the real-time prediction of clique formation to during data collection for experimental verification of a clique classification process discussed in section 8.9.2, or the fitting of multiple aspects to models as described in section 8.8. Relevant research questions are discussed in their respective sections of this chapter.

8.10 Discussion

The research presented in earlier chapters motivates a research effort spanning multiple fields, and has implications for experimental design as well as software engineering and simulation & modelling techniques. Earlier contributions add to the corpus of knowledge in their fields, but also carry

⁷Unfortunately, they remain outwith the scope of this thesis, which focuses on simulation technologies more than it does game design.

Possibly too grandiose? I don't think so really, but I wouldn't want to present it as being more than it is

implications for our understanding of and research practice in those fields. This chapter contributes an investigation into these implications by exploring the possibilities these techniques make possible, and by explaining their potential impact.

In *The structure of scientific revolutions* [Kuh12], Kuhn notes:

“ (...)new inventions of theory [are not] the only scientific events that have revolutionary impact upon the specialists in whose domain they occur. The commitments that govern normal science specify not only what sorts of entities the universe does contain, but also, by implication, those that it does not. It follows, though the point will require extended discussion, that a discovery like that of oxygen or X-rays does not simply add one more item to the population of the scientist’s world. Ultimately it has that effect, but not until the professional community has re-evaluated traditional experimental procedures, altered its conception of entities with which it has long been familiar, and, in the process, shifted the network of theory through which it deals with the world. Scientific fact and theory are not categorically separable, except perhaps within a single tradition of normal-scientific practice. That is why the unexpected discovery is not simply factual in its import and why the scientist’s world is qualitatively transformed as well as quantitatively enriched by fundamental novelties of either fact or theory.”

(...)

The ideas presented in this thesis — in particular aspectually augmented simulations & models and aspect-oriented runtime metaprogramming — are not new “theories” in the relevant fields which contribute any profound or revolutionary new paradigms. However, they do present a new approach in simulation and modelling, and do so through new aspect orientation techniques. There are therefore contributions in both fields in which the “scientific fact and theory” Kuhn refers to are affected: in both fields new research opportunities and experimental practices are made feasible, with implications for both the research facts in the fields (for example, the contribution of the concept of aspect-oriented runtime metaprogramming, for example) and the theory of the fields (for example, the use of new aspect orientation technologies in experimental design).

New capabilities of aspect orientation frameworks are suggested, built, and demonstrated in experimental case studies through PyDySoFu. Their design and the extension of the tooling around these concepts present opportunities for future work as discussed in section 8.5 and in section 8.6, but the implications of their use in software engineering requires further study, as discussed in section 8.1, as

well as in research software engineering specifically, as discussed in section 8.2, section 8.3, section 8.7 and section 8.8; further, the possibilities investigated in this research have implications for research in the practice of research itself, as discussed in section 8.4, section 8.7, section 8.8 and section 8.9.3. For this reason, the work presented in earlier chapters motivates further research in many areas. This thesis contributes not only the exploration of the ideas presented in earlier chapters, but a discussion of their implication for Kuhn’s “scientific fact[s] and theor[ies]” in their relevant fields.

The dataset produced by RPGLite also offers opportunities for further study. Research opportunities presented by the game design and formal methods are explored by Kavanagh [Kav21], but the dataset this thesis helped to contribute also offers opportunities in the mining of player behaviour and the dataset’s use in future aspectually augmented modelling research, particularly if a large-scale data collection study is undertaken. These possibilities were discussed in section 8.9, in particular in section 8.9.1, section 8.9.2, and section 8.9.4.

Maybe go into more detail on what was mentioned in specific cross-referenced sections here? A long list of sections this chapter contains without any extra explanation isn’t very satisfying, even if it’s useful within this framing of exploring what Kuhn meant and how it relates to this thesis.

Chapter 9

Closing Discussion

A chapter rounding out the thesis with a brief discussion.

Bibliography

- [+06] hridayesh rajan hridayesh et al. *nu: towards an aspect-oriented invocation mechanism*. Tech. rep. technical report 414, iowa state university, department of computer science, 2006.
- [AB08] Bourouis Abdelhabib and Belattar Brahim. “JAPROSIM: a Java framework for process interaction discrete event simulation”. In: *Journal of Object Technology* 7.1 (2008), pp. 103–119.
- [ABZ08] AU Aksu, Faruk Belet, and B Zdemir. “Developing aspects for a discrete event simulation system”. In: *Proceedings of the 3rd Turkish Aspect-Oriented Software Development Workshop*. Bilkent University. 2008, pp. 84–93.
- [AGL98] Rakesh Agrawal, Dimitrios Gunopulos, and Frank Leymann. “Mining process models from workflow logs”. In: *Advances in Database Technology—EDBT’98: 6th International Conference on Extending Database Technology Valencia, Spain, March 23–27, 1998 Proceedings* 6. Springer. 1998, pp. 467–483.
- [BB14] Brahim Belattar and Abdelhabib Bourouis. “Yet Another Java Based Discrete-Event Simulation Library.” In: *J. Softw.* 9.1 (2014), pp. 82–88.
- [BH02a] Jason Baker and Wilson Hsieh. “Runtime aspect weaving through metaprogramming”. In: *Proceedings of the 1st international conference on Aspect-oriented software development - AOSD ’02*. ACM Press, 2002. DOI: [10.1145/508386.508396](https://doi.org/10.1145/508386.508396). URL: [https://doi.org/10.1145%2F508386.508396](https://doi.org/10.1145/2F508386.508396).
- [BH02b] Jason Baker and Wilson C. Hsieh. “Maya: Multiple-Dispatch Syntax Extension in Java”. In: *Proceedings of the ACM SIGPLAN 2002 Conference on Programming Language Design and Implementation*. PLDI ’02. Berlin, Germany: Association for Computing Machinery,

- 2002, pp. 270–281. ISBN: 1581134630. DOI: [10.1145/512529.512562](https://doi.org/10.1145/512529.512562). URL: <https://doi.org/10.1145/512529.512562>.
- [Bir99] Colin PD Birch. “A new generalized logistic sigmoid growth equation compared with the Richards growth equation”. In: *Annals of botany* 83.6 (1999), pp. 713–723.
- [Bui14] JCAM Buijs. “Flexible evolutionary algorithms for mining structured process models”. In: *Technische Universiteit Eindhoven* 57 (2014).
- [Cap+09] Claudia Cappelli et al. “An aspect-oriented approach to business process modeling”. In: *Proceedings of the 15th workshop on Early aspects - EA '09*. ACM Press, 2009. DOI: [10.1145/1509825.1509828](https://doi.org/10.1145/1509825.1509828). URL: <https://doi.org/10.1145/1509825.1509828>.
- [CBB13] Meriem Chibani, Brahim Belattar, and Abdelhabib Bourouis. “Toward an aspect-oriented simulation”. In: *International Journal of New Computer Architectures and their Applications (IJNCAA)* 3.1 (2013), pp. 1–10.
- [CBB14] Meriem Chibani, Brahim Belattar, and Abdelhabib Bourouis. “Practical benefits of aspect-oriented programming paradigm in discrete event simulation”. In: *Modelling and Simulation in Engineering* 2014 (2014).
- [CBB19] Meriem Chibani, Brahim Belattar, and Abdelhabib Bourouis. “Using aop in discrete event simulation: A case study with japosim”. In: *International Journal of Applied Mathematics, Computational Science and Systems Engineering* 1 (2019).
- [CCK03] Andy Clement, Adrian Colyer, and Mik Kersten. “Aspect-oriented programming with AJDT”. In: *ECOOP Workshop on Analysis of Aspect-Oriented Software*. Vol. 10. 2003.
- [Cla73] R Lawrence Clark. “LINGUISTIC CONTRIBUTION TO GOTO-LESS PROGRAMMING”. In: *Datamation* 19.12 (1973), pp. 62–63.
- [CM06] Anis Charfi and Mira Mezini. “Aspect-Oriented Workflow Languages”. In: *OTM Conferences*. 2006.
- [CM07] Anis Charfi and Mira Mezini. “Ao4bpel: An aspect-oriented extension to bpel”. In: *World wide web* 10.3 (2007), pp. 309–344.
- [CMM10] Anis Charfi, Heiko Müller, and Mira Mezini. “Aspect-Oriented Business Process Modeling with AO4BPMN”. In: *Modelling Foundations and Applications*. Ed. by Thomas Kühne et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 48–61. ISBN: 978-3-642-13595-8.

- [con16] RafeKettler & github contributors. *magicmethods (A guide to Python's magic methods)*.
<https://github.com/RafeKettler/magicmethods/tree/65cc4a7bf4e72ba18df1ad17a3770>
2011 — 2016.
- [con21] Clarete & github contributors. *ForbiddenFruit*. <https://web.archive.org/web/20210515092416/https://github.com/clarete/forbiddenfruit>. 2021.
- [CS04] Ruzanna Chitchyan and Ian Sommerville. “Comparing dynamic AO systems”. In: 2004.
- [CSS04] Constantinos Constantinides, Therapon Skotiniotis, and Maximilian Stoerzer. “AOP Considered Harmful”. In: *In Proceedings of European Interactive Workshop on Aspects in Software (EIWAS)*. 2004.
- [DBM13] Jan De Houwer, Dermot Barnes-Holmes, and Agnes Moors. “What is learning? On the nature and merits of a functional definition of learning”. In: *Psychonomic bulletin & review* 20 (2013), pp. 631–642.
- [dev21] SimPy developers. *SimPy 4.0.1 documentation*. <https://simpy.readthedocs.io/en/4.0.1/>. 2021.
- [DGM08] Adrian Dozsa, Tudor Girba, and Radu Marinescu. “How lisp systems look different”. In: *2008 12th European Conference on Software Maintenance and Reengineering*. IEEE. 2008, pp. 223–232.
- [Dij68] Edsger W Dijkstra. “Letters to the editor: go to statement considered harmful”. In: *Communications of the ACM* 11.3 (1968), pp. 147–148.
- [Dij82] Edsger W Dijkstra. “On the role of scientific thought”. In: *Selected writings on computing: a personal perspective* (1982), pp. 60–66.
- [Don20] Boudewijn van Dongen. “BPI Challenges: 10 years of real-life datasets”. In: *IEEE Task Force on Process Mining Newsletter #2* (May 14, 2020). URL: <https://www.tf-pm.org/newsletter/newsletter-stream-2-05-2020/bpi-challenges-10-years-of-real-life-datasets> (visited on 05/14/2020).
- [DOR95] DOV DORI. “Object-process Analysis: Maintaining the Balance Between System Structure and Behaviour”. In: *Journal of Logic and Computation* 5.2 (Apr. 1995), pp. 227–249. ISSN: 0955-792X. DOI: [10.1093/logcom/5.2.227](https://doi.org/10.1093/logcom/5.2.227). eprint: <https://academic.oup>.

- com/logcom/article-pdf/5/2/227/6244720/5-2-227.pdf. URL: <https://doi.org/10.1093/logcom/5.2.227>.
- [DR10] Robert Dyer and Hridesh Rajan. "Supporting dynamic aspect-oriented features". In: *ACM Transactions on Software Engineering and Methodology* 20.2 (Aug. 2010), pp. 1–34. DOI: [10.1145/1824760.1824764](https://doi.org/10.1145/1824760.1824764). URL: <https://doi.org/10.1145/1824760.1824764>.
- [Esp04] Miklós Espák. "Aspect-Oriented Programming On Lisp". In: *International Conference on Applied Informatics*. 2004.
- [FBS17] Daniel Friesel, Markus Buschhoff, and Olaf Spinczyk. "Annotations in Operating Systems with Custom AspectC++ Attributes". In: *Proceedings of the 9th Workshop on Programming Languages and Operating Systems*. 2017, pp. 36–42.
- [FFN00] Robert E Filman, Daniel P Friedman, and Peter Norvig. "Aspect-oriented programming is quantification and obliviousness". In: (2000).
- [FT+84] James France, John HM Thornley, et al. *Mathematical models in agriculture*. Butterworths, 1984.
- [Gam+95] Erich Gamma et al. *Design Patterns: Elements of Reusable Object-Oriented Software*. USA: Addison-Wesley Longman Publishing Co., Inc., 1995. ISBN: 0201633612.
- [GK99] László Gulyás and Tamás Kozsik. "The use of aspect-oriented programming in scientific simulations". In: *Proceedings of Sixth Fenno-Ugric Symposium on Software Technology, Estonia*. 1999.
- [Gom15] B. Gompertz. *On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies*. Jan. 1815. DOI: [10.1098/rspl.1815.0271](https://doi.org/10.1098/rspl.1815.0271). URL: <https://doi.org/10.1098/rspl.1815.0271>.
- [GS04] Wasif Gilani and Olaf Spinczyk. "A family of aspect dynamic weavers". In: *Dynamic Aspects Workshop (DAW04)*. 2004.
- [Hie+94] David Hiebeler et al. "The swarm simulation system and individual-based modeling". In: Santa Fe Institute Santa Fe, NM, USA. 1994.

- [Hil+00] Erik Hilsdale et al. "AspectJ: The Language and Support Tools". In: *Addendum to the 2000 Proceedings of the Conference on Object-Oriented Programming, Systems, Languages, and Applications (Addendum)*. OOPSLA '00. Minneapolis, Minnesota, USA: Association for Computing Machinery, 2000, p. 163. ISBN: 1581133073. DOI: [10.1145/367845.368070](https://doi.org/10.1145/367845.368070). URL: <https://doi.org/10.1145/367845.368070>.
- [Ion+09] Tudor B Ionescu et al. "An Aspect-Oriented Approach for Disaster Prevention Simulation Workflows on Supercomputers, Clusters, and Grids". In: *2009 13th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications*. IEEE. 2009, pp. 21–30.
- [JP14] Paul Johannesson and Erik Perjons. *An introduction to design science*. Vol. 10. Springer, 2014.
- [JS07] Chris W Johnson and Christine Shea. "A comparison of the role of degraded modes of operation in the causes of accidents in rail and air traffic management". In: *2007 2nd Institution of Engineering and Technology International Conference on System Safety*. IET. 2007, pp. 89–94.
- [JWO12] Amin Jalali, Petia Wohed, and Chun Ouyang. "Aspect oriented business process modelling with precedence". In: *International Workshop on Business Process Modeling Notation*. Springer. 2012, pp. 23–37.
- [Kav+19] William Kavanagh et al. "Balancing turn-based games with chained strategy generation". In: *IEEE Transactions on Games* 13.2 (2019), pp. 113–122.
- [Kav21] William Kavanagh. "Using probabilistic model checking to balance games". PhD thesis. University of Glasgow, 2021.
- [KDB91] Gregor Kiczales, Jim Des Rivieres, and Daniel G Bobrow. *The art of the metaobject protocol*. MIT press, 1991.
- [Kel08] Stephen Kell. "A survey of practical software adaptation techniques." In: *J. Univers. Comput. Sci.* 14.13 (2008), pp. 2110–2157.
- [KH98] Ralph Keller and Urs Hölzle. "Binary component adaptation". In: *European Conference on Object-Oriented Programming*. Springer. 1998, pp. 307–329.
- [Kic+01] G. Kiczales et al. "An Overview of AspectJ". In: *ECOOP*. 2001.
- [Kic+97] Gregor Kiczales et al. "Aspect-oriented programming". In: *European conference on object-oriented programming*. Springer. 1997, pp. 220–242.

- [KM20] William Kavanagh and Alice Miller. "Gameplay Analysis of Multiplayer Games with Verified Action-Costs". In: *The Computer Games Journal* 10.1-4 (Dec. 2020), pp. 89–110. DOI: [10.1007/s40869-020-00121-5](https://doi.org/10.1007/s40869-020-00121-5). URL: <https://doi.org/10.1007/s40869-020-00121-5>.
- [KM21] William Kavanagh and Alice Miller. "Gameplay analysis of multiplayer games with verified action-costs". In: *The Computer Games Journal* 10 (2021), pp. 89–110.
- [Kuh12] Thomas S Kuhn. *The structure of scientific revolutions*. University of Chicago press, 2012.
- [KWM20] William Kavanagh, Tom Wallis, and Alice Miller. "RPGLite player data and lookup tables". In: (2020).
- [Lac97] Sheldon J Lachman. "Learning is a process: Toward an improved definition of learning". In: *The Journal of psychology* 131.5 (1997), pp. 477–480.
- [LC07] Gary T Leavens and Curtis Clifton. "Multiple concerns in aspect-oriented language design: a language engineering approach to balancing benefits, with examples". In: *Proceedings of the 5th workshop on Software engineering properties of languages and aspect technologies*. 2007, 6–es.
- [Li10] Chen Li. "Mining process model variants: Challenges, techniques, examples". PhD thesis. University of Twente, The Netherlands, 2010.
- [Mac+11] Idarlan Machado et al. "Managing variability in business processes". In: *Proceedings of the 2011 international workshop on Early aspects - EA '11*. ACM Press, 2011. DOI: [10.1145/1960502.1960508](https://doi.org/10.1145/1960502.1960508). URL: <https://doi.org/10.1145/1960502.1960508>.
- [MAG12] Giselle Machado, Vander Alves, and Rohit Gheyi. "Formal Specification and Verification of Well-formedness in Business Process Product Lines". In: *6th Latin American Workshop on Aspect-Oriented Software Development: Advanced Modularization Techniques, LAWASP*. Vol. 12. 2012.
- [Mar94] Robert Martin. "OO design quality metrics". In: *An analysis of dependencies* 12.1 (1994), pp. 151–170.
- [Mat08] Norm Matloff. "Introduction to discrete-event simulation and the simpy language". In: *Davis, CA. Dept of Computer Science. University of California at Davis. Retrieved on August 2.2009* (2008), pp. 1–33.

- [MD11] Richard Millham and Evans Dogbe. "Aspect-Oriented Security and Exception Handling within an Object Oriented System". In: *2011 IEEE 35th Annual Computer Software and Applications Conference Workshops*. July 2011, pp. 321–326. DOI: [10.1109/COMPSACW.2011.60](https://doi.org/10.1109/COMPSACW.2011.60).
- [Mit+17] Alexey A. Mitsyuk et al. "Generating event logs for high-level process models". In: *Simulation Modelling Practice and Theory* 74 (May 2017), pp. 1–16. DOI: [10.1016/j.simpat.2017.01.003](https://doi.org/10.1016/j.simpat.2017.01.003). URL: <https://doi.org/10.1016%2Fj.simpat.2017.01.003>.
- [MK03] Hidehiko Masuhara and Gregor Kiczales. "A modeling framework for aspect-oriented mechanisms". In: *Proceeding ECOOP'03*. 2003.
- [MR02] Katharina Mehner and Awais Rashid. "Towards a standard interface for runtime inspection in AOP environments". In: *Workshop on Tools for Aspect-Oriented Software Development at OOPSLA*. Vol. 2. 2002.
- [Nie77] Jurg Nievergelt. "Information Content of Chess Positions". In: *SIGART Bull.* 62 (Apr. 1977), pp. 13–15. ISSN: 0163-5719. DOI: [10.1145/1045398.1045400](https://doi.org/10.1145/1045398.1045400). URL: <https://doi.org/10.1145/1045398.1045400>.
- [Obj11] Object Management Group (OMG). *Business Process Model and Notation, Version 2.0*. OMG Document Number formal/2011-01-03 (<https://www.omg.org/spec/BPMN/2.0>). 2011.
- [Obj15] Object Management Group (OMG). *Unified Modelling Language, Version 2.5*. OMG Document Number formal/2015-03-01 (<http://www.omg.org/spec/UML/2.5>). 2015.
- [OLe20] Aran O'Leary. "Simulating Human Behaviour in a Micro-epidemic using an Agent-based Model with Fuzzing Aspects". In: *Unpublished masters thesis* (2020).
- [PAG03] Andrei Popovici, Gustavo Alonso, and Thomas Gross. "Just-in-Time Aspects: Efficient Dynamic Weaving for Java". In: *Proceedings of the 2nd International Conference on Aspect-Oriented Software Development*. Association for Computing Machinery, 2003, pp. 100–109.
- [PGA02] Andrei Popovici, Thomas Gross, and Gustavo Alonso. "Dynamic Weaving for Aspect-Oriented Programming". In: *Proceedings of the 1st International Conference on Aspect-Oriented Software Development*. AOSD '02. Enschede, The Netherlands: Association for Computing

- Machinery, 2002, pp. 141–147. ISBN: 158113469X. DOI: [10.1145/508386.508404](https://doi.org/10.1145/508386.508404). URL: <https://doi.org/10.1145/508386.508404>.
- [Pop13] Karl Popper. “The Logic and Evolution of Scientific Theory”. In: Routledge, 2013, pp. 3–22.
- [Pou+15] Asef Pourmasoumi et al. “On Business Process Variants Generation.” In: *CAiSE Forum*. 2015, pp. 179–188.
- [Prz10] Adam Przybylek. “What is Wrong with AOP?” In: *ICSOFT (2)*. Citeseer. 2010, pp. 125–130.
- [RIC59] F. J. RICHARDS. “A Flexible Growth Function for Empirical Use”. In: *Journal of Experimental Botany* 10.2 (June 1959), pp. 290–301. ISSN: 0022-0957. DOI: [10.1093/jxb/10.2.290](https://doi.org/10.1093/jxb/10.2.290). eprint: <https://academic.oup.com/jxb/article-pdf/10/2/290/1411755/10-2-290.pdf>. URL: <https://doi.org/10.1093/jxb/10.2.290>.
- [Ros97] Jarrett Rosenberg. “Some misconceptions about lines of code”. In: *Proceedings fourth international software metrics symposium*. IEEE. 1997, pp. 137–142.
- [SA13] Thomas Stocker and Rafael Accorsi. “Secsy: Security-aware synthesis of process event logs”. In: *Proceedings of the 5th International Workshop on Enterprise Modelling and Information Systems Architectures, St. Gallen, Switzerland*. Citeseer. 2013.
- [SA14] Thomas Stocker and Rafael Accorsi. “SecSy: A Security-oriented Tool for Synthesizing Process Event Logs.” In: *BPM (Demos)*. 2014, p. 71.
- [Seo11] Simon Seow. “OBASHI White Paper”. In: *APMG-International, High Wycombe, UK* (2011).
- [Sil+20] Hercules Sant Ana da Silva Jose et al. “Implementation of Aspect-oriented Business Process Models with Web Services.” In: *Bus. Inf. Syst. Eng.* 62.6 (2020), pp. 561–584.
- [SL07] Olaf Spinczyk and Daniel Lohmann. “The design and implementation of AspectC++”. In: *Knowledge-Based Systems* 20.7 (2007), pp. 636–651.
- [SM14] Ivan Shugurov and Alexey A Mitsyuk. “Generation of a set of event logs with noise”. In: *Proceedings of the Spring/Summer Young Researchers’ Colloquium on Software Engineering*. 8.
 ?????????? ?????????????????? ?????????? ?????????????? ?????? ?????????? ?????????????? ...
 2014.

- [Ste06] Friedrich Steimann. "The paradoxical success of aspect-oriented programming". In: *Proceedings of the 21st annual ACM SIGPLAN conference on Object-oriented programming systems, languages, and applications - OOPSLA '06*. ACM Press, 2006. DOI: [10.1145/1167473.1167514](https://doi.org/10.1145/1167473.1167514). URL: <https://doi.org/10.1145/1167473.1167514>.
- [Sto23] Tim Storer. *Theatre_AG*. https://web.archive.org/web/20230728153002/https://github.com/twsswt/theatre_ag. 2023.
- [Sul01] Gregory T Sullivan. "Aspect-oriented programming using reflection and metaobject protocols". In: *Communications of the ACM* 44.10 (2001), pp. 95–97.
- [SW16] Tim Storer and William Wallis. *Fuzzi-Moss*. <https://web.archive.org/web/20201018121800/https://github.com/twsswt/fuzzi-moss>. 2016.
- [Tha+17] Dai Hai Ton That et al. "Sciunits: Reusable Research Objects". In: *2017 IEEE 13th International Conference on e-Science (e-Science)*. IEEE, Oct. 2017. DOI: [10.1109/escience.2017.51](https://doi.org/10.1109/escience.2017.51). URL: <https://doi.org/10.1109/escience.2017.51>.
- [Van+05] Boudewijn F Van Dongen et al. "The ProM framework: A new era in process mining tool support". In: *Applications and Theory of Petri Nets 2005: 26th International Conference, ICATPN 2005, Miami, USA, June 20-25, 2005. Proceedings* 26. Springer. 2005, pp. 444–454.
- [VD09] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009. ISBN: 1441412697.
- [Ver45] PF Verhulst. "La loi d'accroissement de la population". In: *Nouveaux Memories de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles* 18 (1845), pp. 14–54.
- [Von38] Ludwig Von Bertalanffy. "A quantitative theory of organic growth (inquiries on growth laws. II)". In: *Human biology* 10.2 (1938), pp. 181–213.
- [VVK09] Jussi Vanhatalo, Hagen Völzer, and Jana Koehler. "The refined process structure tree". In: *Data & Knowledge Engineering* 68.9 (2009), pp. 793–818.
- [VW04] Wil MP Van der Aalst and Anton JMM Weijters. "Process mining: a research agenda". In: *Computers in industry* 53.3 (2004), pp. 231–244.
- [Wal+20] William Wallis et al. "Designing a Mobile Game to Generate Player Data-Lessons Learned". In: (2020).

- [Wal+21] William Wallis et al. "Designing a mobile game to generate player data - lessons learned THIS IS A DUPLICATE FIXME THIS IS A DUPLICATE FIXMETHIS IS A DUPLICATE FIXMETHIS IS A DUPLICATE FIXME THIS IS A DUPLICATE FIXME THIS IS A DUPLICATE FIXME THIS IS A DUPLICATE FIXME THIS IS A DUPLICATE FIXME **Duplicate citation to clean up**". In: *CoRR abs/2101.07144* (2021). arXiv: [2101.07144](https://arxiv.org/abs/2101.07144). URL: <https://arxiv.org/abs/2101.07144>.
- [WC10] Paul Wallis and Fergus Cloughley. *The OBASHI Methodology*. The Stationery Office, 2010.
- [WJ97] AR Werker and KW Jaggard. "Modelling asymmetrical growth curves that rise and then fall: Applications to foliage dynamics of sugar beet (*Beta vulgaris*L.)" In: *Annals of Botany* 79.6 (1997), pp. 657–665.
- [WS18a] Tom Wallis and Tim Storer. "Modelling realistic user behaviour in information systems simulations as fuzzing aspects". In: *International Conference on Advanced Information Systems Engineering*. Springer. 2018, pp. 254–268.
- [WS18b] Tom Wallis and Tim Storer. "Process Fuzzing as an Approach to Genetic Programming". In: *Proceedings of the SICSA Workshop on Reasoning, Learning and Explainability*. Ed. by Kyle Martin, Nirmalie Wiratunga, and Leslie S. Smith. Vol. 2151. CEUR Workshop Proceedings. Aberdeen, UK: CEUR-WS.org, June 2018. URL: http://ceur-ws.org/Vol-2151/Paper_S3.pdf.
- [WS18c] Tom Wallis and Tim Storer. *PyDySoFu*. <https://github.com/twsswt/pydysofu>. 2018.
- [WS23] Tom Wallis and Tim Storer. *ASP*. <https://web.archive.org/web/20230730164214/https://github.com/probablytom/asp>. 2023.
- [YC79] Edward Yourdon and Larry L Constantine. "Structured design. Fundamentals of a discipline of computer program and systems design". In: *Englewood Cliffs: Yourdon Press* (1979).