Approval

Name: Ye Wang
Degree: Master of Science
Title: Tools and Tasks: How Designers Make and Use Tools for Design Explorations

Examining Committee:
Chair: Wolfgang Stuerzlinger
Professor

Halil Erhan
Senior Supervisor
Associate Professor

Brian Fisher
Supervisor
Associate Professor

Robert Woodbury
External Examiner
University Professor
SIAT

Date Defended: August 21, 2015
Ethics Statement

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

a. human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

b. advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

c. as a co-investigator, collaborator or research assistant in a research project approved in advance,

or

d. as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Simon Fraser University Library
Burnaby, British Columbia, Canada

update Spring 2010
Abstract

Design exploration is a complex creative activity that has common characteristics across disciplines. The lack of a unified theory of design exploration poses challenges for designing Computational Design Tools (CDTs), especially with respect to how designers make and use exploration tools in different task environments. Through three related studies, I aim to understand how designers act in the exploration processes. Study 1 investigates conceptual sketching in webpage design explorations. Study 2 comprises interviews of architecture practitioners about their use of exploration tools. Study 3 is a lab study of the architecture schematic design exploration process given a large number of CDT generated alternatives. Experimental interactive visualizations were built to assist with design exploration and data analysis. From these studies, we conjecture that overloaded design exploration presents a task environment distinctly different from the non-overloaded case. This thesis concludes with system interaction design suggestions for future CDTs from our findings.

Keywords: design exploration; design cognition; computational design tools; cognitive styles; parametric tools; exploring design alternatives
To my father.
Thank you for fighting it.

...and THIS is not your fault.
Really.
Acknowledgements

There are times that I thought I’d never make it during my time at SIAT. It has been the most difficult, but also the most rewarding few years for me personally and intellectually. I sincerely thank everyone who helped me to learn, grow and survive. I wish I could name everybody here and thank each one of you for completing this wonderful journey with me, but I can only name a few.

THANK YOU, THESIS COMMITTEE! I would like to express my deep gratitude towards Dr. Halil Erhan for being my senior supervisor, and getting me through the program. A big thank you to Dr. Brian Fisher for introducing me to SIAT and generously letting me sit in his own lab meetings. Also I’d like to thank Dr. Rob Woodbury for taking the time to be my external examiner. Lastly, appreciations to Dr. Bernhard Riecke, who regrettably was not able to attend my defense, for all his advices over the years.

THANK YOU, EDITORS! I would like to thank Jacob Freiberg, and editors from fiverr.com for their help in proofreading my thesis.

THANK YOU, FOR YOUR INSPIRATIONS! Other than the guidance from my thesis committee, I am very fortunate to be influenced by many great professors and researchers at SIAT. I’d like to thank Linda Kaastra, Chris Shaw, Steve DiPoala, Chantal Gibson, all members of the Computational Design Lab, SCIENCE Lab and attendees of VA reading group for their instructions, inspirations and feedback. Also, I would like to thank all the friends that I have made during SIAT, especially Naghmi Shireen, Rodolfo Sanchez, Jelena Popovic, Mengting Sun, Chao Feng and Xin Tong. Lastly, I’d like to thank all authors that I have cited in this thesis, including those generously granted me image copyright permissions.

THANK YOU, PARTICIPANTS! I would like to thank all study participants, especially industry professionals in Study 2 for taking the time to help out. (Sorry, can’t name you because of anonymity rules.) Study 1 expert panel - thank you as well!

THANK YOU, SIAT STAFF! Special shout-outs to Gordon Pritchard and Larry Soo for their professional IT support. Also I’d like to say thank you to Tiffany Taylor, Joyce Black and Desiree Nazareth for their administrative support.

THANK YOU, FAMILY & FRIENDS! I would like to thank both of my parents for setting great examples for me during the most difficult times. And to all my friends that helped me through challenging personal times, especially Kaka.

THANK YOU, YOU. You know who you are. Enough said.
# Table of Contents

Approval ii
Ethics Statement iii
Abstract iv
Dedication v
Acknowledgements vi
Table of Contents vii
List of Tables xi
List of Figures xii
List of Acroynms xvii

1 Introduction 1
1.1 Research Goal & Questions ........................................ 3
1.2 Motivation ........................................................... 4
1.3 Thesis Structure .................................................... 7

2 Related Work 9
2.1 The D-Word (design) ................................................ 9
2.2 Studying The Design Process .................................... 10
2.3 Design (Space) Exploration ...................................... 13
2.4 Studying Alternatives .............................................. 17
2.4.1 How many alternatives? ....................................... 17
2.4.2 But why more alternatives? ................................... 19
2.5 Variables in Design Studies ..................................... 25
2.5.1 Cognitive Styles ................................................ 25
2.5.2 (Ir)rational Decision Making ................................. 30
2.5.3 Domain of Design .............................................. 31
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1.2 Coding Scheme Development Attempt</td>
<td>84</td>
</tr>
<tr>
<td>5.2 Methods</td>
<td>84</td>
</tr>
<tr>
<td>5.2.1 Participants</td>
<td>84</td>
</tr>
<tr>
<td>5.2.2 Task</td>
<td>84</td>
</tr>
<tr>
<td>5.2.3 Procedures</td>
<td>85</td>
</tr>
<tr>
<td>5.2.4 Data Collection and Analysis</td>
<td>87</td>
</tr>
<tr>
<td>5.2.5 Design Process Visualization</td>
<td>89</td>
</tr>
<tr>
<td>5.3 Results</td>
<td>90</td>
</tr>
<tr>
<td>5.3.1 Self Descriptions of Design Exploration Process</td>
<td>100</td>
</tr>
<tr>
<td>5.3.2 Challenges in Narrowing Down Overloaded Alternatives</td>
<td>104</td>
</tr>
<tr>
<td>5.3.3 How Many Alternatives is Too Many?</td>
<td>104</td>
</tr>
<tr>
<td>5.3.4 Example Analysis with Interactive Visualization (with P1 data)</td>
<td>106</td>
</tr>
<tr>
<td>5.4 Discussion</td>
<td>113</td>
</tr>
<tr>
<td>5.5 Conclusion</td>
<td>116</td>
</tr>
<tr>
<td>6 Overloaded Design Exploration in CDTs</td>
<td>117</td>
</tr>
<tr>
<td>7 Lessons Learned For Designing Computational Design Tools (CDTs)</td>
<td>122</td>
</tr>
<tr>
<td>7.1 Support Modular Alternative Explorations</td>
<td>122</td>
</tr>
<tr>
<td>7.2 Recognize and Support Different Roles in Design Team</td>
<td>123</td>
</tr>
<tr>
<td>7.3 Support Different Numbers of Alternatives</td>
<td>125</td>
</tr>
<tr>
<td>7.4 Encourage Integration of Software Practices into Design Practices</td>
<td>126</td>
</tr>
<tr>
<td>7.5 Enable Effortless Transition Between Digital and Physical Artifacts</td>
<td>127</td>
</tr>
<tr>
<td>7.6 Build Trust Between Human and Machine.</td>
<td>129</td>
</tr>
<tr>
<td>Do What Each Other is Best At, Together.</td>
<td>129</td>
</tr>
<tr>
<td>7.6.1 Constant reinforcements of performance merits to users</td>
<td>129</td>
</tr>
<tr>
<td>7.6.2 Make procedures more transparent</td>
<td>130</td>
</tr>
<tr>
<td>7.6.3 Involve user decisions as much as possible</td>
<td>130</td>
</tr>
<tr>
<td>7.6.4 Do things the users way</td>
<td>130</td>
</tr>
<tr>
<td>7.6.5 Know what the users are the better at machines.</td>
<td>132</td>
</tr>
<tr>
<td>8 Conclusion</td>
<td>133</td>
</tr>
<tr>
<td>Bibliography</td>
<td>135</td>
</tr>
<tr>
<td>Appendix A Study 1: Post-Task Questionnaire</td>
<td>144</td>
</tr>
<tr>
<td>Appendix B Semi-Auto Transcription Tool</td>
<td>148</td>
</tr>
<tr>
<td>Appendix C Case Study: Attempts at Parametric Reduction</td>
<td>151</td>
</tr>
<tr>
<td>C.1 Introduction</td>
<td>151</td>
</tr>
</tbody>
</table>
List of Tables

Table 2.1 Description of Dreyfus’ Model of Skill Acquisition (Five Stage Cognitive Model) (Cheetham and Chivers, 2005) ........................................ 34
Table 3.1 Means of Responses. (E): expert-rated. (S): self-reported. ............. 43
Table 3.2 Distribution of Initial Task Action Responses. .......................... 47
Table 4.1 Summary of the interviewed participants (P) and firm background. .... 59
Table 4.2 Interview script. Clarification and follow-up questions are not included. . . 60
Table 4.3 Different processes classified by the person(s) involved in executing the CAD typewriter role. .......................................................... 65
Table 4.4 Process characteristics of Scenario A: Museum Project ................ 79
Table 4.5 Process characteristics of Scenario B: High-rise Residential Project ... 80
Table 4.6 Process characteristics of Scenario C: Cultural Centre Project .......... 81
Table 4.7 Process characteristics of Scenario D: Mix-Use Development .......... 82
Table 5.1 Design task description given to participants. ............................ 87
Table 5.2 Number of alternatives reported by participants for different conditions. ... 105
Table 5.3 Participants’ estimates of number of alternatives on cards viewed (Post-Q5). ** P0 data unavailable. .................................................. 106
List of Figures

Figure 1.1 Thesis Research Overview: This thesis is attempting to complete part of the identify-propose-build cycle. Solid lines represent parts of the cycle covered by the thesis, dotted lines represent parts not covered in detail. Circled numbers are corresponding chapter numbers. 8

Figure 2.1 Maher’s model of design as co-evolution 14
Figure 2.2 Maher’s model of design as exploration 15
Figure 2.3 Munzner’s illustration on effective design space exploration for vis designers 20
Figure 2.4 Extending Munzner’s illustration on effective design space exploration for vis designers 20
Figure 2.5 Shah et al’s metric of variety 21
Figure 2.6 Example from Shah et al’s work illustrating the genealogy tree resulted from a golf ball collector design activity. 22
Figure 2.7 Different types of encodings of existing variety metrics adopted from Fuge et al 23

Figure 3.1 Correlations of expert panel scores 42
Figure 3.2 Distributions of joint scores through expert panel consensus for vagueness, novelty and usability (n = 18). Error bars showing standard error. 42
Figure 3.3 Correlations between self-rated and expert-rated scores for novelty and usability. 42
Figure 3.4 Means of Expert-rated Vagueness scores (Y Axis) vs. Style of Design Alternatives Generation (X Axis). Significant main effect, \( F(3, 17) = 7.47, p = 0.003^* \) 44
Figure 3.5 Means of Expert-rated Novelty scores (Y Axis) vs. Style of Design Alternatives Generation (X Axis). Significant main effect, \( F(3, 17) = 4.81, p = 0.017^* \) 44
Figure 3.6 Means of Self-Reported Usability scores (Y Axis) vs. Style of Design Alternatives Generation (X Axis). Significant main effect, \( F(3, 17) = 5.17, p = 0.013^* \) 44
Figure 3.7 Significant difference in expert-rated vagueness scores for no alternatives 45
Figure 3.8 Significant difference in expert-rated novelty scores for no alternatives 45
Figure 3.9 Means of Self-Judged Expertise scores (Y Axis) vs. Self-Reported Experience in Task Domain (X Axis). Significant main effect, $F(2, 17) = 22.60, p < 0.0001$.
(a) no experience, $n = 9, M = 3.67, SE = 0.44$.
(b) only unpaid experience, $n = 5, M = 6.80, SE = 0.49$
(c) with paid experience, $n = 4, M = 7.75, SE = 0.25$

Figure 3.10 Means of Expert-Rated Vagueness scores (Y Axis) vs. Occupation (X Axis). designers: $n = 7, M = 6.21, SE = 0.92$.
other occupations: $n = 7, M = 4.79, SE = 0.64$.
science / engineering disciplines: $n = 4, M = 3.00, SE = 0.89$.

Figure 3.11 Variables that showed significant correlations with each other.

Figure 3.12 Heatmap of distribution of subject style of alternative generation by occupation and experience levels described in Section 3.2.5.

Figure 3.13 Different alternative generation processes observed.

Figure 3.14 Sample sketches from P1, showing 2 alternatives out of the total three produced. The visual tour alternative on the left aims to replicate the experience of shopping at a physical store location, by allowing the users to navigate through the merchandises by floors and brand display sections. The generic inventory alternative on the right has a conventional shopping site view.

Figure 3.15 Sample sketches from P13, showing all 4 alternatives produced. The subject extensively explored different site layouts and user interactions, such as animation flows, zooming behaviour, and drag-and-drop actions.

Figure 3.16 Sample sketches showing detailed sketching by a single solution participant (P6). It contains detailed components of the web page views.

Figure 3.17 Sample sketches showing detailed design by single solution participant (P10). It includes both the web page views and site structures.

Figure 3.18 Sample sketches showing four very vague and abstract alternatives by P18, exploring site layouts.

Figure 3.19 Sample sketches from P16 based on a user scenario approach.

Figure 3.20 Sample sketches from 7 subjects, who all produced a single solution. Some have multiple sheets of paper illustrating different parts of the same idea.

Figure 4.1 Example of Sketch to CAD artifact transition.

Figure 4.2 Example of sketch on CAD to new CAD artifact transition.

Figure 4.3 Example of CAD to new sketch artifact transition.

Figure 4.4 Examples of 2D: non-digital renderings.

Figure 4.5 Example of 2D drawings with printed CAD drawings + non-digital media, and its creation process.
Figure 4.6 Examples of 2D renderings with detailed CAD drawings. .......................... 72
Figure 4.7 Examples of photo-realistic 3D architectural rendering. .......................... 73
Figure 4.8 Examples of physical 3D models built during the design process. ........... 74
Figure 5.1 Description of Participant Background Experience. .............................. 85
Figure 5.2 Sample generations from the set used in the study. ............................... 86
Figure 5.3 Observational Study Room Setup. ....................................................... 88
Figure 5.4 Description of Participant Background Experience. .............................. 89
Figure 5.5 A partial view from the spreadsheet showing all coded events from P1’s videos. ......................................................................................................................... 89
Figure 5.6 An overview screenshot of the design process visualization in basic mode ................................. 90
Figure 5.7 Zoom-in views of various interaction details of the process visualization. .. 91
Figure 5.8 Examples of two design alternatives visualized clearly showing different level of interactions from the participant during the study session. ........... 91
Figure 5.9 Observed cycle for design exploration with a large number of computer generated alternatives. Numbers in the graph represent steps in the cycle, corresponding to detailed descriptions in the text. ............................... 92
Figure 5.10 P9 and P5’s final deliverables for design exploration task, with both based on a single alternative. ................................................................. 94
Figure 5.11 P6 and P7’s final deliverables for design exploration task. Both based on inspirations from multiple alternatives seen, but combined with the designer’s own elements, result in a final design that did not resemble any of the alternatives in the generated set. ................................. 94
Figure 5.12 P4’s final deliverables for design exploration task. Two alternatives produced with different views (top, front, back, side) are based on inspirations from multiple alternatives seen, and combined with designer’s own elements, resulting in a final design that did not resemble any of the alternatives in the generated set. ................................. 94
Figure 5.13 P2 viewing alternatives one set at the time, attempting to develop criteria. 96
Figure 5.14 P2 viewing alternatives one at a time using only non-physical recorded verbal criteria. ................................................................. 96
Figure 5.15 P2 Design Exploration Process. ......................................................... 97
Figure 5.16 P2 Design Exploration Process (Last Cycle: 58 min to 1 h 58 min). ....... 98
Figure 5.17 P4 Design Exploration Process .......................................................... 99
Figure 5.18 Interim sketches produced by P4 during the exploration process. Final deliverable sketches are shown in Figure 5.12. ................................. 100
Figure 5.19 P6 design exploration process for Cycle 1 and Cycle 2. Lettered bullets are elimination criteria used or newly developed (in darker red) during that time of exploration. ................................. 101
Figure 5.20  Example of an abstract visual self description of the design exploration process: P0 (pilot subject). ................................. 102
Figure 5.21  Examples of procedural textual self description of the design exploration process: P4 and P9. ................................. 103
Figure 5.22  Examples of Detailed Visual Self Description of the Design Exploration Process: P1, P7 and P8. ................................. 103
Figure 5.23  Observed design exploration cycles in P1’s process, adapted from generic description of steps in Figure 5.9. ................................. 107
Figure 5.24  Illustration of P1’s temporal design patterns adapted from a screenshot of the interactive visualization. ................................. 108
Figure 5.25  P1’s criteria development events adapted from scaled-down screenshot. ................................. 108
Figure 5.26  P1 Cycle 1 - Initial speculation and strategy development ................................. 108
Figure 5.27  P1 Cycle 1 - Developed more explicit criteria on post-it notes. ................................. 109
Figure 5.28  P1’s “ruler” for evaluating the number of floors, with a self-made post-it note 110
Figure 5.29  P1 Cycle 2 - Followed the same strategy and criteria developed in Cycle 1, and added more criteria during the process. Left: Start of Cycle 2. Right: Nearly at the end of Cycle 2. ................................. 110
Figure 5.30  P1 Design Process (Cycle 2 Detailed View) ................................. 111
Figure 5.31  P1 Design Process (Cycle 3 Detailed View) ................................. 112
Figure 5.32  P1 Cycle 3 - Continuing same strategy and criteria as Cycle 2, showing samples of the post-its [stilt buildings] or [earthquake proof] categories evaluated as two parallel concepts. ................................. 112
Figure 5.33  P1 Design Process (Cycle 4 Detailed View) ................................. 113
Figure 5.34  P1 Cycle 4 - Final state of the design exploration process. ................................. 113
Figure 5.35  Examples of two alternatives that had been discarded initially, but were reverted back to being temporarily accepted, and eventually lasted till the end of the session. ................................. 114
Figure 5.36  Example of an alternative that has been through temporarily accepted, discarded, to multiple interactions and eventually still discarded. ................................. 115
Figure 5.37  Example of P3’s exploration process, in which he uses vertical placements to indicate groupings of discarded alternatives. ................................. 115
Figure 6.1  Sequential use of criteria during design exploration. Users move alternatives through each criteria to filter out some of them based on selection or elimination. ................................. 119
Figure 6.2  Criteria used as a “set” of alternatives. It represents a collection of alternatives that are grouped based on whether they passed or failed a certain criterion. Sometimes these groups are used as an exemplar collection to direct the design exploration by the “do’s” pile and the “don’ts” pile. . . 119
Figure 6.3  Re-arrangement of alternatives’ ranking within each criteria during design exploration. .......................................................... 120

Figure 6.4  Criteria used as a "tag" in design exploration. When an alternative is dragged and dropped from one criteria to another, it is "tagged" with both criteria, instead of leaving the first one. This way the alternative can belong to different criteria simultaneously. .................................................. 120

Figure 7.1  Example of two parallel modular alternative explorations (i.e. windows, exterior colours) and how they can be represented and merged with a version-control structure. .................................................. 123

Figure 7.2  CDT’s should support easy transition between artifacts during any phase of the design, including design exploration. ................................. 128
# List of Acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFS</td>
<td>Breadth-First Search</td>
</tr>
<tr>
<td>CAD</td>
<td>Computer-Aided Design</td>
</tr>
<tr>
<td>CD</td>
<td>Construction Document (phase in architectural design)</td>
</tr>
<tr>
<td>CDT</td>
<td>Computational Design Tools</td>
</tr>
<tr>
<td>DD</td>
<td>Design Development (phase in architectural design)</td>
</tr>
<tr>
<td>DFS</td>
<td>Depth-First Search</td>
</tr>
<tr>
<td>DSE</td>
<td>Design Space Exploration</td>
</tr>
<tr>
<td>G-CDT</td>
<td>Generative Computational Design Tools</td>
</tr>
<tr>
<td>IS</td>
<td>Information Systems</td>
</tr>
<tr>
<td>NDE</td>
<td>Overloaded Design Exploration</td>
</tr>
<tr>
<td>ODE</td>
<td>Non-Overloaded Design Exploration</td>
</tr>
<tr>
<td>SD</td>
<td>Schematic Design (phase in architectural design)</td>
</tr>
<tr>
<td>VA</td>
<td>Visual Analytics</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Design is a creative endeavor, of which the process is not well understood and is still under much debate. However, there is agreement that working with design alternatives is an essential part of a design activity, despite much disagreements on how designers work with alternatives. Advances in computing hardware and software are involving computational tools in more and more stages of the design process. Technologies such as machine learning and parallel computing are gradually changing some traditional design processes, especially when brainstorming, exploring, and evaluating design alternatives. The complex and multi-faceted nature of design makes it challenging to build computational tools to support designers to harvest this new power that was not available to the generations of masters before them. Computational tools might be new friends to the artists.

One of the biggest problem with computers is that they are not designers. The great American graphic designer Milton Glaser (known for the ‘I ♥ NY’ logo) once said, "Computers are to design as microwaves are to cooking." I agree with this clear analogy in many ways. Design is an art like cooking, and computers are just one of many tools assisting designers, just as microwaves complements the chefs’ cooking process alongside many other tools like an oven or a stove. Microwaves can be a great invention on their own, but they are not good at everything. A microwaves can bake potatoes much faster than an oven, but if you try to hard boil an egg with a microwave, you will probably have a major explosion. Less known to many, the mechanics of the microwave actually enables it to produce perfect quiches, banana bread, salmon, meat loaf, etc., but why don’t chefs use it to do serious cooking as much as a stove or an oven? Let me make some informal speculations about the microwave and try to convince you that our problem in computational design is exactly the same as that of the microwave:

1. **Microwaved food easily gets “too dry” compared to conventionally cooked food.** A microwave’s cooking profile is very different from conventional heating elements in the oven or stove that is radiating water molecules at a much lower frequency. In microwaves, water molecules are more likely to evaporate, and heat less evenly. In design, generative computational tools are also working in a different way than traditional approaches. Some designers even have strong emotional responses towards computationally generated solutions, and
consider them ‘too dry’ and lacking of the human touch to be interesting. Most generative computational tools are based on enumerating through a set of parameters in a digital design representation. The alternative solution set has a different profile than a human generated one: 1) many more alternatives available, 2) much more concrete, 3) may or may not have enough differences between individual alternatives as judged by the designer.

2. **Large differences across models and manufacturers.** Different microwaves can behave very differently due to varying power output, internal engineering differences including turning plate design, etc. Warming up a cup of milk to 180F takes a different amount of time on different microwaves, while traditional ovens give consistent results when specifying 375F for 20 minutes. In computational design, the similar problem of inconsistent behaviour for generating alternatives across various tools or even within the same tool makes it challenging for designers to understand and utilize the tools most effectively. In some cases, it might be much easier for designers to use manual techniques for alternative generation. Some argue the results are more understandable and are deemed to be more quality-controlled by the designers themselves.

3. **Small optimal cooking time window and difficulty in monitoring food state once cooking has started.** Due to microwave’s high energy, the amount of time required to cook the same amount of food using a microwave is only a fraction of the time it takes a conventional oven/stove. For example, cooking a potato using a microwave calls for more precision on time setting than does an oven. Cooking for one minute below or over the optimum window can yield either an over-cooked or an under-cooked potato, and the action of opening the door to check on the potato will interrupt the heating process. Compounding with manufacturing differences among microwaves, it becomes even more difficult to follow any microwave recipes than traditional recipes. While using computational design tools to generate alternatives, designers often faced the same difficult problem of controlling the generated set. It is difficult for the designer to control the process (e.g. pure combinatoric or evolutionary parametric generation) during generation of alternatives. Although some of the systems attempt to incorporate control into the iterative design process, the user-interaction mostly happens after an action for generation instead of giving real-time feedback. Designers have to wait until the generation process is finished before they can interact, and then regenerate again, if the results are not satisfying. This interaction model is different from the traditional reflective practice of designers, where they constantly reflect and modify their designs during the cycle. To designers, the optimal window for a computer-generated set of alternative solutions is very small, just as with the optimal cooking time of the microwave. The alternatives set often either falls short for not being containing enough quality solutions or for having too many solutions for the designers to go through.

Now you see there are many reasons why traditionally trained master chefs assert that the microwave is not helpful for serious cooking. For the same reasons, master designers, like Milton
Glaser, tend to claim the same thing; generative computational design tools are not helpful for serious design.

If you really think about it, microwaves have the exact same capabilities as any traditional cooking device, namely that you can cook at varying power levels and for a flexible amount of time. Theoretically, it should do all the things an oven or stove can do. **If I were a microwave marketer, what would I do to make people use it more?**

- Make users realize microwave cooking has different properties than traditional cooking.

- Improve the ability to monitor and interact with food while cooking without interrupting the heating process. For example, add infrared sensors to detect food temperatures. This enables the user to get real time feedback, much like the experience of “watching the stove”, and give more control over their cooking strategy. Another added feature could be adding motorized components that can be controlled by the user when the microwave door is closed, mimicking the ability to flip or stir the food while it’s cooking.

- Add more preset function buttons, specifically user-tested and fine-tuned, to each model for common cooking needs.

Following this microwave analogy, it is easier for me to describe the research goal of this thesis. We believe that Computational Design Tools (CDT) have the potential to achieve the same or better quality of design as traditional design processes, at least within some task domains (if not all). We wish to propose interface recommendations allowing the designers to interact with the alternatives with more efficacy and efficiency during design exploration. CDT’s have the potential to deliver design solutions with less demand on time and expertise.

### 1.1 Research Goal & Questions

With three incremental studies with different focus and methodology, this thesis attempts to achieve the following research goal:

- Outline interface design ideas building on study findings to maximize capabilities of computational design tools for design space exploration.

The two research questions that need to be answered to achieve this goal through the studies are:

- Learn possible behaviour differences in design processes among designers in traditional design as well as computational design practices, in particular, in generating and evaluating alternatives during design space explorations.

- Identify what aspects of computational design tools designers might better assist designers compared to traditional design tools, and what aspects are under-utilized in design space explorations and why.
Some concrete contributions (intentional and unintentional) that resulted from this research include:

- Exploratory investigations into how designers generate and evaluate alternatives during the conceptual design phase in different context e.g. with and without CDT’s, with varying numbers of alternatives.

- A survey (small sample) on design tools use in architectural design practice.

- A demonstration of using visual analytics for joint human-machine design decision-making mixed-initiatives by applying theories and algorithms.

- A demonstration of using visual analytics for video data analysis based on partial protocol analysis.

- A discussion of directions for future confirmatory quantitative studies based on interesting qualitative observations from this exploratory work.

We investigated through site interviews and experimental user studies to understand these questions:

- How are expert designers different than novices in terms of handling alternatives? [Study 1: Mix Methods Observational Study, Controlled Lab Setting] (Chapter 3).

- How are current computational design tools are used in the architecture industry, and what are they lacking in supporting the designers to handle more alternatives? [Study 2: On-site Interviews, Architecture Office Setting] (Chapter 4).

- How do designers handle large numbers of alternatives in a real design task? [Study 3: Mix Methods Observational Study, Controlled Lab Setting] (Chapter 5).

1.2 Motivation

Design is a complex activity. Different schools can’t even agree on the definition of “design” itself. There are heavy debates among design researchers on how design processes are carried out, how design should be taught and what really makes a designer a master designer. Research in this area is challenging due to the difficulties in evaluating design outcomes and observing cognitive design processes that are only implicit in the master minds. This thesis is attempting to generalize our findings mainly in architectural design and web interface design fields with the intention to form a bigger picture of design practice. Much literature suggests cognitive patterns in design are shared across different fields (see Section 2.1, 2.2). One common traits among design experts in different fields is that they tend to generate more design alternatives during the conceptual design phase. The phenomenon is widely described as breadth-first, then depth-first, as opposed to depth-first, which
occurs only in novice extremes Woodbury and Burrow (2006b). I attempt to frame my research carefully within established design theories and will address our concerns and limitations caused by challenges in design research.

Beyond design research, computational design research has special battles of its own. In particular, questions of how much of a role and which roles should computational systems play in the process of creative design are heavily debated. The traditional view in design is represented by Schön and other master designers, which asserts that design is a reflective practice that has a vast component of subjectivity, style, implicit expert knowledge that can only be gained through doing Schön (1983). In this view, CDT have limited roles in creating designs, but mostly with representing or coding design ideas. Users have dialogues with the tools, whether it is a pen or a system, and the tacit flow, including the sensory perceiving of how the pen maneuverous on paper, is part of the conversation. Users use the tools through an interactive conversation, rather than using it to create designs for themWinograd (1996). On the other extreme to the traditional views, Simon takes a stand that design is no different than a problem-solving task, and it is a simple recall of knowledge and information synthesis. Simon (1996) He believes we think of as expert judgment and inspiration are actually subconscious recall of information, which may include involuntary pattern recognition, such as seen in master chess players. We believe he would argue that understanding the cognitive processes during design can enable us to create design tools that partially solve design problems without much interference ((Simon, 1996, page 136)). The majority of computational design researchers sit in between Schon and Simon’s view, believing some aspects of creative design can be replaced by computational tools, and others cannot. The difference among computational design researchers are the questions of what can one day be replaced, should be replaced, and how designers behave differently when such actions are replaced.

In this thesis, we take the general stance that designers should be involved in screening the alternatives in the iterative design cycle. In this context, screening means selecting and eliminating generated alternatives, or changing generation flow as needed to seek design opportunities. It is not an optimization procedure, but rather part of design, to explore diverse yet plausible alternatives. We envision a designer’s power of creativity will be unleashed during this interactive process to achieve full impact with the computational powers of creation (that may present unexpected solutions that feed a designer’s inspiration). The interactive and iterative cycle can cause crucial changes in the final design outcome. We founded our stand in Chapter 2 with existing literature. We believe that using generative computational tools can increase the variety (thereby, increasing the “breadth”) in initial conceptual design phases for designers. “Breath” is considered a desired quality in expert design practice. It is accepted that weeding through all the generated alternatives (easily in the thousands, if not millions combinatorial space) is the big challenge. Before the designers get involved in the screening process, the computational system should be capable of cutting the space down to a manageable scale for human cognitive limits. The system should also present such alternatives in a way that is easy for designers to perceive and interact with the alternatives in the screening process. The ultimate research goal of our research program is to eventually implement and test the impact
of such generative computational tools. This thesis aims to understand how designers interact with alternatives, and to gain insights into how to design the next generation of computational tools.

During the initial stage of this research, there were two burning questions on my mind. The first one was: how exactly do novice and expert designers differ in their interactions with alternatives? As a novice designer myself I had a strong motivation to uncover how to become an expert designer. However, more importantly, I felt if we wanted to build a decent computational design tool, it should not only be effective in supporting experts, but should also build successful expert patterns into the interactive system to assist novices with learning expert practices and grow faster. An example of such systems are software development IDEs. The code-assist features in Eclipse or Visual Studio constantly remind or force novice developers to conform to good coding practices. I believe that, the best tools were the ones that enable novices to perform at the same level as the experts in minimum time and without the necessary knowledge. In other words, with help from such system, the novice is still a novice. It is not a lazy way for the novice to take the shortcut to develop necessary expert skills, but for novices it is instead a kick start on more complicated tasks that their current skill-level prevents them from accomplishing. For example, a GPS system instantly allows Kate, who’s directionally challenged, to drive to her friend’s wedding as accurately as Tom, who knows the city inside out. Without the GPS, Kate still won’t be able to drive to the correct place herself. I would like to make design recommendations to a computational tool that supports experts and empowers the novices at the same time. In the context of the GPS example, we want to find ways to support Tom without boring him with too many details that he already knows (e.g. quick glance modes before driving), as well as help Kate navigate as efficiently as Tom. As the work progressed, it became less important to identify what differentiates experts and novices per se, but rather to investigate which design processes lead to better outcomes despite when used by experts or novices. In order to better understand different design practices and how to best support all designers, I ran the first observational study described in Chapter 3. It investigates differences in working with alternatives generation and evaluation in conceptual web page design. This study also provided an interesting cross-reference to CDT design in hind-sight.

Then I moved on to my second burning question. We observed that generative design tools are becoming increasingly matured in the architecture industry. However, are practicing designers relying on these tools as a core technology to assist with design exploration? If yes, how do they use them? If not, why? The motivation behind inquiring about the current status of the industry on the use of existing generative computational tools was to determine if we are not barking at the wrong tree. Many sub-questions came from this line of concerns, e.g. How many alternatives do designers normally create? With or without the tool? Do designers feel there is a need to generate a large number of alternatives at all? How many alternatives are too many for designers? How do we know the large numbers of alternatives generation tools are not just another computer scientists’ creature that fitted their geek egos? Sure, we can spit out all these different designs, but are they useful for the designers in the way they work (prime example: all of Windows 8’s powerful but impractical new features)? The interviews conducted at architecture firms in Vancouver and New York in Chapter
4 (Study 2) seek answers to these questions. It was a relief to find that we have enough reasons to believe using generative computational tools at large scales might be useful for designers to enrich their design exploration phase. Valuable insights from practitioners were gathered regarding generative computational tools. There was a vast difference across the spectrum in terms of design experience, computational tool literacy, area of practice, firm size, etc.

After promising confirmations from the first two studies, we decided there was a need to conduct an observational study involving expert designers performing a design exploration task with a large number of generated alternatives (Study 3). Chapter 5 attempts to make a small step towards filling the current gap in the literature on design exploration and generative computational tools. Although the algorithms for generating and optimizing design alternatives are moving forward at a steady pace, the interface and interaction techniques of the generative tools often falls short of supporting effective exploration. By observing designers behaviour during conceptual design alternatives exploration for an architectural design task among 1000 computer-generated alternatives, we hope to discover behavioural patterns in the design processes involving a large number of generated solutions.

Due to lack of existing literature of design exploration behaviour in generative computational tools, we are in need of a classification of designers’ behaviours and their activity patterns for generative design tool making. Developing a reproducible coding scheme is the first step towards being able to perform confirmatory quantitative studies for any hypothesis we have. Summarizing and triangulating results from all three studies, we hope this thesis will provide a foundation for a classification for cognitive activities in design exploration with a large number of generated alternatives. We also hope the qualitative observations from this thesis will be the start of a line of research in the future.

Finally, I strive to produce a set of design recommendations for developing generative design support tools that may be applied in multiple design domains for exploring alternative conceptual solutions in early phases of design.

1.3 Thesis Structure

The structure of this thesis follows a design science research method for inquiry Hevner et al. (2004), with an emphasis on behavioural science investigations with a goal of suggesting systems design guidelines (Figure 1.1).

I begin by reviewing design literature in Chapter 2 to define design and design (space) exploration. Then I review existing design studies related to how designers generate and evaluate alternatives, and psychology literature in cognitive styles and decision making. In Chapter 3 to Chapter 5, I present three studies that probe how designers explore alternatives in different settings. In the first study (Chapter 3), I explore different design behaviours and processes during a conceptual web page design task in people with various levels of design experience. The second study (Chapter 4) is an interview study that focuses more on how designers in the architecture industry behave in
design explorations, and how they work with CDTs. With some insights from Study 1 and Study 2, we designed the last study (Chapter 5) to investigate design exploration behaviours with computer generated alternatives for a schematic architectural design task. Finally, in Chapter 8, I propose a set of user experience (UX) design recommendations for CDT development based on findings from all three studies.
Chapter 2

Related Work

2.1 The D-Word (design)

Design is a complex and multi-faceted creative discipline. The question of “What is Design?” seems to be a very difficult one to answer even among designers (Winograd, 1996, Chapter 8). Two major tensioned design paradigms are 1) design as a problem-solving activity, and 2) design as a reflective process. Simon (1996) takes the first approach, while Schön (1983) represents the second one. In many design literature, the two paradigms are often positioned as opposing theories of design with one more focused on (bounded) rationality and the other on intuition. However, a close reading of works of Simon (1996) and works of Schön reveals surprising similarities in their understandings of the design process (Schön, 1988; Schön and Wiggins, 1992; Winograd, 1996, Chapter 9). Soo Meng (2009) has discussed many valuable points in how the misinterpretation of Simon’s work in general design literature may have compounded over the years, starting with Schön and Cross.

I see the two theories as non-contradicting, rather represented at different level of abstractions. Their unit of study is no different as well on their definition of a task environment. Simon’s theory has strong basis in formalizing concepts and processes with a computer science flavour, a top-down approach. Schön’s theory has strong basis in protocol studies with a professional practitioner flavour, a bottom-up approach. Both recognizes the complexity of the design process. Both agrees that design is “partly empirical”, and “partly tacit”. Both agrees on the difficulty in “ill-defined” problems, and that new goals can emerge during the design process and feeds back to the design cycle. It is perhaps because Simon’s theory is higher on the abstraction scale than Schön, that became the source of mis-reading among designers. I believe the two theories indeed can meet in the middle, drawing from Simon’s line of work for formal definitions and descriptions of process, and drawing from Schön’s line of work for real-life examples. This thesis subscribes to the common ground between Simon and Schön as the foundation of the following discussions in the domain.

Design can be an overloaded word. In this context, we use Simon’s definition that encompasses the broader sense of design:
'Everyone designs who devises courses of action aimed at changing existing situations into preferred ones. The intellectual activity that produces material artifacts is no different fundamentally from the one that prescribes remedies for a sick patient or the one that devises a new sales plan for a company or a social welfare policy for a state....Schools of engineering, as well as schools of architecture, business, education, law, and medicine, are all centrally concerned with the process of design.” (Simon, 1996, page 111)

By this definition, design processes (at least at an abstract-level) share many common characteristics. Researchers including Akin (2001) and Gero (2006) verifies this in their studies. Hence, we assume that researching findings in design processes may be generalized across domains, e.g. architectural design, graphic design, web design, user experience design, interactive visualization design. Even in disciplines that traditionally are not considered as design (in the narrow aesthetics sense), such as software design, vaccine design, nursing protocol design, course curriculum design, experimental design etc. The growing hype about design thinking in both popular and academic design literature is postulating that there is a set of cognitive activities that is specific to design processes across domains.

This thesis references literature from different domains to study behaviour differences in design processes, and also hypothesizing that the results from our studies in web design (Chapter 3) and architectural design (Chapter 4 and 5) may be applicable across domains.

2.2 Studying The Design Process

Despite disagreements on the details on design processes, in discussions about design thinking, there is a generally accepted seven-step process: define, research, ideate, prototype, choose, implement, and learn (Moser, 2013, page 90) All parties agree that the process is not linear, can occur simultaneously and can be repeated. From Simon (1996)’s “Generate-And-Test” to McKim (1980)’s “Express-And-Test”, to Schön (1983)’s “reflection-in-action & reflection-on-action”, to Roozenburg (1995)’s “basic design cycle”, to Brown (2009)’s “empathy-creativity-rationality”, the central theme of the seven-step process and the nonlinear nature of design is shared among all variants. Although by definite of the processes, in some variants multiple steps are combined into one, and/or given a different label.

Most of these design theories are based on analytical formalization, semi-formal observations, anecdotal stories and protocol analysis of low number of participants. Some may argue that these studies are not as reproducible as some of the other more quantified studies. How to study design processes to confirm or reject the above theories is a quite controversial subject. The juxtaposition of science and design have produced many confusing terms containing different combinations of these two words: design science, scientific design, the science of design, design science research, design as science. The confusion comes from different connotations of different words, as well as the same word being used differently by different people.
In this thesis, we are using Cross (2001)’s version of term differentiations to further clarify our use of terms. Four main aspects or views about the science-design relationship are:

1. Design requires application of scientific knowledge from various domains (e.g. materials science, engineering science, behavioral science)

2. Design of scientific procedures is required in scientific or engineering problems with no established methods

3. Design itself is a scientific practice that has systematic and logical laws and rules that govern its activities like natural sciences such as physics

4. Design as an activity can be investigated through scientific methods

The term scientific design is claimed to be used “uncontroversially” (Cross, 2001) in the design community for meaning in bullet #1, to reflect modern industrialized design utilizing scientific advances as opposed to pre-industrialized craft-oriented design practices. However, in science and engineering domains, it is used often to represent meanings in bullet #2 (Beitz, 1994; Rinaldi et al., 2009; Xue et al., 2008).

The term design as science is usually referring to meaning in bullet #3, to contrast with design as discipline. This is the main conceptual difference between Schön and Simon’s views, despite what I think are actually agreements in between the two views in explaining and observing design processes.

Design science is said to be first coined by Buckminster Fuller (Cross, 2001) to represent meanings in bullet #3 and used interchangeably with design as science in some design literature. However, as the design research community shifting towards a more empirical approach to study design, the word is taking more meaning of bullet #4 or both. Gero’s line of work in protocol analysis (FBS ontology framework) is exemplar work of the current design science research community that prescribes both bullet #3 & #4 (Gero and Kannengiesser, 2004; Gero and Neill, 1998). Hevner and Chatterjee (2010) also agree with both bullet #3 & #4 with a focus on Information Systems (IS) design, but using the word design-science research in a different sense:

“The fundamental principle of design science research is that knowledge and understanding of a design problem and its solution are acquired in the building and application of an artifact.” (Hevner and Chatterjee, 2010, page 5)

Hevner and Chatterjee (2010) claim that design science research must rely on building and applying artifacts (i.e. information systems), and differentiates the studies on design phenomenon without building and testing IS artifacts are behaviour science research. It was argued behaviour-science research and design-science interacts with each other, and are complementary paradigms in the same IS research framework to provide a more complete understanding of design problems. In this view, many design literature such as Gero’s work on investigating sketching behaviour in design is considered behaviour-science research instead of design-science research.
The science of design in Simon’s work explicitly takes on the meaning of bullet #3 in a chapter in his seminal book *The Sciences of the Artificial* with the exact title (Simon, 1996, page 118). However, Cross interprets the word as bullet #4’s meaning, and further states that recognizing (bullet #4) design can be studied scientifically is independent of whether (bullet #3) design is a scientific activity or not (Cross, 2001).

To avoid confusions on terms, we clarify that we agree with all four bullets about the design-science relationship described above, and consider the word design science research or design science encompasses all these four aspects. We use these terms interchangeably, and avoid using other terms discussed without disambiguating the exact meaning.

We believe the study of design processes is included in design science research as we see it:

- can and should be investigated through scientific methods
- can provide insights into eventually being able to describe and teach design as a scientific practice
- can produce results that can be generalized across all design domain as defined in the last subsection

Although parts of this study would be classified as behaviour-science research by Hevner and Chatterjee due to the lack of building and evaluation of IS artifacts, we believe the investigations around cognitive behaviours in design tasks in the absence of IS artifacts is a valuable step before developing such IS for further investigation. We insist on including this type of work in our classification of design science research, as it is not only useful, but in our opinion, necessary to know how people design before you build IS to support people design.

After clarifying and situating our position on the relationship between design and science, there is a need to reaffirm our respect for the basic principles of science relating to falsifiability throughout this work. We believe what fundamentally drives scientific discoveries are two principles:

1. Science requires all proposed theories to be falsifiable.

2. Science allows any conjectures to exist as theories, until someone is able to refute it empirically.


“It is easy to obtain confirmations, or verifications, for nearly every theory - if we look for confirmations. Confirmations should count only if they are the result of risky predictions; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory - an event which would have refuted the theory.”
It is the cycle of constantly establishing “risky predictions” and attempting to refute them that led to more robust revisions to existing theories and establishment of new theories. I believe the predictions that stucked around as recognized theories are those still falsifiable, but failed to be refuted for a while.

This thesis on investigating design processes is based on the above principles of science in an attempt to make “risky leaps” in proposing falsifiable hypotheses through observations. We argue these proposals should be allowed to stand, however wild they are, until someone refutes them.

2.3 Design (Space) Exploration

It is commonly accepted that there is or should be an exploration phase in design processes across all domains. This process is commonly referred to as design exploration or design space exploration. Sometimes it is tied to a temporal frame or phase in the design process. For example, in design literature it is common to state the authors would like to investigate “design exploration during conceptual design phase”, but leaves the actual definition of “design exploration” unexplained, assuming the audience understands it. Like design and design science, the word design (space) exploration is also overloaded, and should be clarified for the purpose of our discussion.

In engineering design disciplines, such as hardware interface design, the term can be defined to differentiate the design process with the implementation process, such as this one:

“Design Space Exploration (DSE) refers to the activity of exploring design alternatives prior to implementation.” (Kang et al., 2011)

This definition considers “design space exploration” the entire phase before deciding on the final implementation plan. It is a clear-cut definition, works well in engineering domains where design and implementation are separate processes. But it is difficult to apply across domain, as in some design domains (e.g. graphic design), the design itself is the implementation. By this definition, the entire design process will be exploration. Not what we are looking for.

In computational design research, “design exploration” is often defined in relation to “design optimization”. According to Simon (1996), the design process can be abstracted as a problem-solving activity, which can further abstract to a search process. Simon does recognize the complex nature of design, problem of ill-defined problems and changing goals, and believes the abstraction of design as a search process is accurate, just may be that current technologies in computation and neuroscience have not advanced to uncover all the mechanisms of such as search yet. Along this line of reasoning, computational design research from different domains (such as Mechanical Engineering and Architecture Design) are agreeing on the differences between design exploration and design optimization when both are modeled as search processes:

“The essential difference between design optimization and design exploration is the method for characterizing the outcome. Design optimization strategies have two dis-
distinct parts; formulate and converge. Here it is assumed that the problem can be formulated before the search and convergence begins. Design exploration strategies, on the other hand, are based on the belief that the problem formulation evolves during the process of searching and converging, thus ultimately leading to a more informed optimal solution. In this way, design exploration is both divergent and convergent.” (Mattson, 2014)

The interpretation is that “optimization” requires a pre-defined unchanging goal and search space, while “exploration” has an unclear and changing goal and search space. Only during the exploration search process, its goal becomes more refined through iterative changes of the search space.

Maher (2000) models the design process as a co-evolutionary process by differentiating the three formalizations of design:

- **Design as search**: “Design can be formalised as search when the goals of the design are well-defined before search commences and the focuses of design are not changed until a solution is found.”

- **Design as exploration**: “Design becomes exploration when the focus of the design changes as the process continues. A focus, derived from the requirement space, determines which solution in a solution space is considered best.”

- **Design as co-evolution**: “Design becomes co-evolution when the focus of the design can change, and the requirements space and solution space change through mutual interaction. The focus of the search is based on the requirements when searching the solution space, and based on the solutions when searching the requirement space.” (Figure 2.1)

![Figure 2.1: Maher (2000)’s model of design as co-evolution, with kind permission from Springer Science and Business Media](image)

In Maher’s definition, “design as exploration” seems to be differentiating itself with “design as co-evolution” by its unchanging problem space and solution space. On each iteration of the “design
as exploration”, a different focus of the problem space produced a difference focus on the solution space, as illustrated by Maher in Figure 2.2. The final selection from “design as exploration” is selected from solutions focused from all iterations.

![Figure 2.2: Maher (2000)'s model of design as exploration, with kind permission from Springer Science and Business Media](image)

In this thesis, we are adopting Maher’s “design as co-evolution” model to describe the design (space) exploration process, which includes iteration cycles of the problem space and solution space informing each other to proceed with the design process. Here, design space is including both the problem space and solution space. This view is rooted in ideas pioneered by Simon (1996) in his earlier works.

A wide body of research in creativity (e.g. design fixation, ideation, collaboration) covers the terms concept space and idea space (Howard et al., 2008; Segers et al., 2005; Verhaegen et al., 2013). Some researchers use the term concept and idea interchangeably, but we make the distinction between the two terms as articulated by Srinivasan and Chakrabarti (2010). A concept is represented by a set of ideas at different abstraction levels. An idea can appear in various concepts. In the context of generative design tools, we are mostly concerned with full concept generations when we refer to alternative generation. We would consider both concept space and idea space are encompassed by the solution space, and can map onto each other formally. Clevenger and Haymaker (2011) identified the lack of agreement on term definitions in design research, and proposed a set of “performance-based design definitions” to precisely coin design spaces into the objective space, alternative space, impact space and value space. We see this as a detailed breakdown of design space, where by our definitions, the objective space is part of the problem space, the alternative space is the equivalent of the concept space, while the impact space and value space are relationships between the problem space and the solution space.
In computational design research, the focus of word “design space exploration” is on representing and organizing the space to assist the designers to explore:

“Design space exploration is the idea that computers can be used to help designers by representing many designs, organizing them in a network structure that forms the space, and by assisting designers to explore this space: that is, to make new designs and to move among previously discovered designs in the network.” (Stouffs, 2006)

Seminal work by Woodbury and Burrow (2006b) on “design space” falls into this category of definition. Most of Woodbury et al’s work are focused on the formalizing representations of the “network structure of designs” with typed feature structures. This is immensely valuable in building computational design programs, however, outside of the scope of this thesis. We interpret Woodbury et al’s work as being mostly at the designer’s “action-level”, and mostly focused on formal representations of the solution space - of which Woodbury et al argue includes the problem space (i.e. requirements). And we differentiate our work as with more focus on the interactions between “problem space” and “solution space” during the design exploration phase, at a generic decision-making level, which is applicable to all levels of design exploration.

There are many levels of design exploration as noted, but we believe the common decision-making strategies and patterns will appear in all levels of design explorations. As Flemming (2006) points out, there are at least three levels: computational level, symbol level, and knowledge level. Flemming refers to Woodbury’s work being at the computational level for representing the designer’s “action layer”, whereas there is an additional “task layer” or “program layer” at the symbol level, which is different from other layers in the abstract knowledge level above it. Woodbury and Burrow (2006a) argues there is no need to “single” out particular layers, as there would be arbitrarily many layers of abstraction. Each of the layer ultimately calls down to its lower layers that eventually are represented by the lowest computational layer. We recognize both Flemming and Woodbury’s views, and situate our work more in the abstract “knowledge level” of strategies, which must be observed at “action” level for scientific studies, moreover, must be represented at and translates into the computational level.

We take the approach the “problem space” is difficult in most real life design problems across domains, especially due to the often “ill-defined” and implicit nature. CDTs are still not yet fully capable of formally representing the “problem space” though many attempted to address the challenges head-on, e.g. Erhan (2003)’s RaBBiT. However, CDTs are already capable enough to generate vast “solution spaces” with the input of human designers. In mixed-initiative systems, most of the higher level problem spaces are difficult to model formally within the system. We thereby are more interested in interaction between the “problem space” and the “solution space” during the design space exploration. We hope to make design recommendations with a focus on improving problem-solution space interactions based on existing solution space representations.

Design exploration are sometimes defined by its goals, such as these definitions:
“The main purpose of design exploration is to identify the relationship between the performance of the product (maximum stress, mass, fluid flow, velocities, etc.) and the design variables (dimensions, loads, material properties, etc.)” (Canonsburg, 2013)

“Exploration is the process of generating and evaluating design alternatives that normally would not be considered [by designers]” (Navinchandra, 1991)

To recognize all the definitions discussed above, when we are referring to “design exploration” or “design space exploration” or “design (space) exploration”, we mean:

• a co-evolution process of the problem-space and solution-space

• a process that is different from design optimization, and often used together with design optimization to accomplish a design goal

• a process that contains a component for verifying against problem requirements such performance, but also seeks for surprises

For the scope of this thesis, when we discuss “behaviours” during “design (space) exploration”, we are referring to the cognitive activities that is shared across design domains for generating and evaluating alternatives at different levels of abstraction.

2.4 Studying Alternatives

From the definitions design (space) exploration in the previous section, we can see that “generating and evaluating alternatives” is one of the main activities during this process. The study of alternatives has a long-standing history in design literature across domains. Many studies have been done in comparing novice and expert designers in different strategies in generating and evaluating design alternatives, but there are many disagreements in findings.

In the next sections, we focus on reviewing the literature around the two questions of our interest:

• How many alternatives is too many for designers to deal with?

• What are the different strategies do designers use to generate and evaluate alternatives?

We also briefly discuss other factors that influences the answers to these questions but are outside the scope of this thesis.

2.4.1 How many alternatives?

With the goal of building effective design (space) exploration computational tools, a key question is how many alternatives is too many for the designers to handle. It can be further broken down into:
• does this max number of alternatives cognitive limit have a big range among individual designers?
• is this number different across different design domains?
• is this number different when different tools are used?
• is exploring more alternatives better than exploring less alternatives for designers?

Ultimately, we are trying to build tools to support interactions with the optimal number of design alternatives, and we are unsure whether there is such an optimal number or range.

We attempted to review literature relating to human cognition limits on attending to the number of alternatives. Miller’s famous magic seven experiments indicates that humans can hold 7+/−2 items in the short memory. Independent of Miller’s findings, other studies investigating the effectiveness of human response to number of alternatives seem to coincide with this number. Cox III (1980) suggests that the optimal number of response variables on a scale (such as those given on questionnaires in user studies) are best kept between 5-9.

In the automotive industry, it is found that the optimal number of alternatives to be developed in parallel are 1-3 depending on linearity of cost structure (Schäfer and Sorensen, 2010). Akin observed expert designers generate an average 2.5 - 4 alternatives in both Engineering and Architects (Akin, 2001). Chai et al. (2015) finds that experts spends less time on generating alternatives comparing to novices, in contradictory of classic design literature. To extrapolate on results from Chai et al. (2015, Table 6), experts generate an average of 3.4 design concepts, while 1st year students and 3rd year university students generate 8 and 2.5 design concepts respectively.

We also looked into decision-making literature outside of the design domain. Iyengar and Lepper (2000) studied the effect of choice-overload in a series of experiments that compare choice behaviors between controlled (6 choices) vs. overloaded choices (24 or 30 choices). They argue that in tasks of choosing chocolate, jam, essay writing, having more choice alternatives is correlated to less likely to make a decision to carry out the action (e.g. to buy a jam or write the essay). Although being very inspiring study, unfortunately the results from her three studies did not converge. The results demonstrate that more alternatives is correlated with more enjoyment and engagement of the selection, but also less satisfaction (more regrets) in the final decision. In pension consumer behaviour research, it is also shown that people are more likely to buy a 401k plan when a handful of alternatives are given as opposed to 10 or more options (Mitchell and Utkus, 2004, Chapter 5).

Others have shown that different representations or tools can affect the number of max alternatives optimal for designers. Wujec (2013) argued that using individual note cards increase the number of nodes users can consider concurrently than one piece of static paper.

To correspond to the few questions raised around “how many alternatives is too many” with a review of existing literature seems to suggest the following starting point for our research:

• the human short term memory and attention cognitive limits seems to be at 7 +/- 2 under various context, even with individual differences
• the range of design alternatives generated by human designers (accounting for individual experiences) in various design studies are generally (well) under 20, with averages 2.5 to 4.

• choice overloading experiments in consumer economics show more alternatives (10+) presented by sellers / creators may correlate with more enjoyment during the choosing experience, but less likely to purchase / choose and more likely to regret after an actual choice comparing to when less options are offered.

All of the above suggest human may have a somewhat limited processing ability for alternatives, regardless whether the alternatives are generated by the decision maker or not. Although there is no agreed optimal number, but the number is at the scale of less than 10, rather than thousands or millions or alternatives, which is what the generative design tools are capable today. In this thesis, we try to investigate for ourselves whether this human cognitive processing ability is really limited at the scale of <10 alternatives. If yes, how could we build tools to respect this limit and augment human information processing capabilities in design? If no, what is the range limits of alternatives that can be processed by human?

It is worth mentioning that all of the studies mentioned above do not have an effective alternatives exploration mechanism for the decision maker. Like argued in the introduction section with microwave interface analogy, we believe building an effective interface might be the booster to human cognitive limits to process more alternatives.

2.4.2 But why more alternatives?

Now, the next question is why do we want to support a broad design space exploration that includes more alternatives?

A great illustration (Figure 2.3) by Munzner (2015, page 13) shows why viz designers should consider a broader space when designing information visualizations. The outer square indicates a space with all possible solutions in the world. Then the biggest circle indicates the “known space”, containing solutions the designer is aware of. The smaller “consideration space” is what the designer actively explores, and the even smaller “proposal space” is what the designer investigates in detail and finally narrows in on the final one solution. In all design domains, initially considering a smaller solution space may exclude better solutions. Good design practice asks for explorations of a broader solution space that includes more potential “good solutions” than just “OK solutions”.

Does “more” alternatives in pure quantity terms guarantees “bigger” consideration space? Probably not (See Figure 2.4). More in quantity does not guarantee good solution, as the consideration space may be restricted to more poor alternatives.

Then, what exactly do we mean by “more” alternatives? We actually mean more alternatives that with overall variety. In other words, more diversity in the alternatives in the design space, more differences among the alternatives.

A widely cited variety metric is by Shah and Vargas-Hernandez (2003). The authors developed a set of metrics to measure the ideation process: quantity, quality, novelty and variety. The variety
metric is for evaluating the overall differences within a set of alternatives, as opposed to evaluate differences between two alternatives. It requires compilation of a genealogy tree with four levels, to code the set of alternatives based on different functions presented in each alternative. The four levels, namely physical principles, working principles, embodiment and detail, are weighted 10, 6, 3, 1 respectively from top-down, signifying more abstracted branching differences gives more variety than differences in lower-level details.

Shah and Vargas-Hernandez (2003)'s work is very inspiring theoretically as a quantitative metric for variety. However, the authors were not very clear on the practicalities of how the functions constructing the tree are derived from the set of alternatives initially. In their example for variety calculations, the task was to design “a semi-autonomous device to collect golf balls from a playing field and bring them to a storage area”. The two functions “pick/collection balls and transport/maneuverer
Figure 2.5: Shah et al’s metric of variety. Where $M_3$ is the variety metric of the set of alternatives; $m$ is the total number of functions; $S_k$ is the score for level $k$ (which is 10, 6, 3, 1 respectively); $b_k$ is the no. of branches at level $k$; $n$ is the total no. of alternatives (Shah and Vargas-Hernandez, 2003).

device” were explicitly described in the paper, but unsure whether were presented to the subjects or not along with the design task. In the novelty section of the paper, Shah et al described two ways of constructing the function/features list: priori or posteriori. It would be reasonable to assume the authors may take similar approaches for the variety metric. The priori approach requires the researchers to come up with a list of functions before the experiment, and the posteriori approach gathers all alternatives and derives a list of functions from them. The priori method ensures the metric is not biased by the results, but runs the risk of ecological validity in more complex design tasks (e.g. design a smart phone), where concrete functionalities are not clearly defined initially but evolves with the process (e.g. selfie functionality, location-tracking after being lost or stolen). Pre-defined functionality-list even by the designers themselves can be limiting to be used as for the variety metric. The posteriori method, on the other hand, is difficult to be done without biases. Furthermore, there were little descriptions in Shah et al’s work in how to define each of the four levels, and when to count alternatives as a separate branch. (e.g. in Shah et al’s example, rolling by wheels and rolling by threads are counted under the same node “rolling” at physical level, but two different nodes at the working level. See Figure 2.6.) However, one may see this as two different physical nodes as wheels and threads are two physical things. These ambiguities in the metric definition can affect the consistency in the evaluations, and can perhaps only be countered by a blind inter-rater review stage before data analysis to construct a functions list, and genealogy tree structures.

Shah et al’s metric of variety is not suitable for the purpose of our study, for the following reasons:

1. **Too domain specific.** The four levels in the genealogy tree of functions are meaningful in examples given by Shah et al (mechanical engineering), however, may be difficult to be applied in other design domains (as per our definition for design in earlier sections). For example, one of the levels are physical principle – some of design domains may be completely lacking this level (e.g. mobile application UI design, policy design), but have their own domain specific high-level variety abstractions (e.g. menu interaction models: navigation drawer, multi-level drop-down etc. )

2. **Too vague to be followed in practice.** Although design studies refer to Shah et al’s creativity metrics, the variety metric is used less than quantity and novelty. In studies that explicitly uses Shah et al’s variety metric (Okudan et al., 2010; Song and Agogino, 2004; Sun et al., 2014; Weir et al., 2005), metric evaluation procedures were either not described in detail, or
Figure 2.6: Example from Shah et al’s work illustrating the genealogy tree resulted from a golf ball collector design activity. Reprinted from Design Studies, 24(2), Shah, JJ. Smith, S.M., Vargas-Hernandez, N., Metrics for measuring ideation effectiveness, 129, ©2013, with permission from Elsevier (Shah and Vargas-Hernandez, 2003).

described as being very different from one another (i.e. priori or posteriori; how functions at each level are defined; whether multiple raters are used etc). A good number of papers reference Shah et al’s variety metric only as a background, but proceeds to use different ways of evaluating variety that are easier to execute, for example:

- **bin sorting**: sort all alternatives into as many bins as needed, where each bin contain a group of similar alternatives. The variety score of a set of alternatives is calculated by the number of bins the alternatives are in divided by the total number of bins (Linsey et al., 2011).

- **feature counting**: attribute features (also called functions, subfunctions, classifications etc.) to each alternative. Features can be mutually exclusive at a single level of abstraction, or at multiple levels. The count of all features presented in a set of alternatives is used as the variety metric of the set. A morphological table is sometimes used (Grantham et al., 2013; Richardson III et al., 2011; Worinkeng et al., 2013)
• **individual variety scoring:** each individual alternative is given an absolute variety score based on its properties, as opposed to a metric based on the entire set’s properties (Sipilä and Pertula, 2006). This is actually departing from Shah’s definition of variety.

3. **Too expensive to perform.** Existing evaluation variety reviewed above all requires human expert(s) for manual scoring. Commonly involving the following steps:

(a) establish a coding structure (e.g. linear, hierarchical and graphical representations), and assign scores to each node on the structure. e.g. Shah et al.’s genealogy tree is hierarchical, and morphological tables can be considered a linear look-up. Defining the levels and the tables is establishing the coding structure.

(b) fit the set of alternatives into the structure.

(c) compute the final metric based on scoring rules and values read off from the structure

This process is unrealistic in the context of evaluating design alternatives generated by computational means due to the sheer number of alternatives generated. Fuge et al. (2013) attempts to automate variety evaluation via machine learning algorithms (addresses the arbitrary weighting’s problem), but still cannot overcome the manual coding in step (a).

![Figure 2.7: Different types of encodings of existing variety metrics adopted from Fuge et al. (2013), reproduced with kind permissions from Mark Fuge.](image)

4. **Too arbitrary.** As Verhaegen et al. (2013) pointed out, the weights are arbitrarily chosen in Shah et al and other variety metrics along this line of research (Nelson et al., 2009; Sarkar and Chakrabarti, 2011). To address this issue, Verhaegen proposed a level-specific metric for variety to avoid the problem of assigning arbitrary weights, as well as a Herfindahl index to represent the uniformness of the distribution. This approach opens doors for looking at variety from a different angle, it gives enriched profiles of the set variety rather than a single number Shah et al provides. However, it is still difficult compare two sets of alternatives in terms of overall variety - even more difficult than the single number metric, especially in larger sets as is with most computational generative cases.
Since all existing metrics for variety are not ideal for studying generative design tools, we must face the challenge of developing a workable variety metric in our studies, by adapting and revising the current metrics or perhaps reframe this problem in a different manner.

To accomplish our goal to build better generative computational design tool, the variety metric is crucial (but not limited) to the following aspects:

1. to assist the designers in automatically screening the large generated alternatives set for a reasonable set with high variety
2. to assist in evaluating generative design tools for the effectiveness of different algorithms or interaction / visual designs in terms of output alternatives variety level
3. to better understand the correlations of variety against other design variables, for example, expertise level, design strategies, design output etc.

For the complex nature of design, the case by case, person to person differences, we believe there might be an approach in measuring variety that involves more human decision for each individual project and each individual designer, but at the same time making the process more automatic and objective. This may sound contradicting, let’s break it down.

Generative computational design tools is capable of generating thousands or billions. we need some kind of function to map the parameter space to a personalized function of variety for the specific project. For example, in one project, the design exploration process calls for maximum visual difference in building structures, while in another project, it may ask for maximum difference in building heights. In most projects, the designers may not have a clue on how to start, and just wants to see a variety of solutions in all dimensions. This need for flexibility in the variety metric itself is poorly addressed by existing metrics. Even Fuge et al. (2013)’s metric that utilizes machine learning only offers flexibility in determining weights to a list of predefined features’ variety score, but does not offer flexibility in the list of features themselves. We think it might make sense to provide ways for the designers to define what the variety means in the initial phase of the exploration, and continues to enable them to modify the metric if they wish. This is what we mean by “more human involvement”. What we mean by “more automatic” is that the definition of variety is an abstract mapping directly onto the parametric space. Therefore, the designer does not need to evaluate all the alternatives one by one, rather only a small set of alternatives to provide the metric definition and its upcoming revisions. In Shah and Vargas-Hernandez (2003)’s metric, an evaluator is required to construct the genealogy trees of alternatives for each function. In Appendix C, we discuss in detail how this is accomplished through a case study, and how this method can be generalized in other generative design domains.

There are many challenges in attempting to approach the variety metric in this way. Our approach definitely have its own weaknesses, and is far from being a solid model. But we hope this work is nudging the direction of variety evaluation slightly, and will inspire further design studies in eventually coming up with a solid evaluation metric for variety.
2.5 Variables in Design Studies

This section briefly discusses the commonly studied variables in design literature and the correlations / interactions among them, while focusing on the ones that we are particularly interested in examining. One of the goals of this thesis is to classifying different cognitive strategies and patterns in design decision making with generated alternatives. All other variables in discussion below can be confounding variables to our study, and are potentially fruitful research areas to be investigated on top of findings of this work.

2.5.1 Cognitive Styles

In classic psychology literature, the concept of cognitive styles seems to go in and out of style since its first popular era in the 1950s. Cognitive styles, defined as “people’s characteristic and typically preferred modes of processing information”, have been studied and applied in education and occupational context widely (Sternberg and Grigorenko, 1997). There is an implied connotation that these styles are intrinsic to the person, relatively stable overtime, (may change according to the environment, but may also be difficult to change) - and therefore educators / employers should take different approaches to adapt to various types of students / employees for best outcomes (Cassidy, 2004; Sternberg and Grigorenko, 1997). Historical basic cognitive science research in cognitive styles focused on identifying cognitive style dimensions with two opposing poles (Kozhevnikov, 2007, page 468). Though individual differences are demonstrated, lack of agreement between data trials, as well as lack of attempt to integrate all the dimensions. Basic research on cognitive styles declined, since the segmented findings in different dimensions were impractical in offering a coherent theory to explain individual differences or predict behaviours (Kozhevnikov et al., 2014). Despite the declined interest in basic cognitive research, applied research on cognitive styles in more complex tasks trended in the 1970s, and has been particularly well received in three areas: business decision-making styles, education learning / teaching styles, personal styles. Most research have a two-level model, identifies with both the perception level addressed by previous basic research and the level of complex cognitive activities such as decision making and problem solving.

Defining Individual Cognitive Styles

During our literature review in both basic and applied cognitive styles, we found there are a lot of confusion and ineffective dispersion of research results caused by, 1) cross-labeling the same style for different names, 2) using the same term for different dimensions, 3) intentionally disagreement or unintentionally mis-labeling of terms, 4) non-mutually exclusive (overlapping) style definitions. We agree with Kozhevnikov et al. (2014) that a robust unifying taxonomy is urgently needed for applied fields to be able to utilize decades of cognitive styles research effectively. Some of the commonly seen cognitive styles are listed here.
- **Field Dependence-Independence**  
  *The tendency to rely on information or structure in the prevailing visual field.*  
  This is first well studied cognitive style dimension, dating back to the 1940s by Witkin. Common tests include the Rod and Frame Test (RFT), Embedded Figures Test (EFT), Body Adjustment Test (Clewley et al., 2010; Witkin and Goodenough, 1981).

- **Wholistic-Analytic**  
  *The tendency to understand concepts wholistically about the entire system versus analytically making sense about parts to the systems.*  
  This cognitive style was initially widely studied to explain differences between artists / musicians (wholistic) and scientists / engineers (analytic). It has also been used commonly as a name for a cognitive style family that includes dimensions like “field dependence - independence” and “impulsivity-reflectivity” (Goodale et al., 2014; Pitta-Pantazi and Christou, 2008; Riding and Cheema, 1991). Sometimes it is used as the style family name, but referring to a specific (narrow) cognitive style.

- **Convergent-Divergent Thinking**  
  *The tendency to generate multiple solutions or deduce to a single solution to a problem.*  
  This cognitive style is first coined by Guilford in the Aptitudes Research Project (Guilford, 1967). In recent design studies in academia and popular design methodology books, these terms have extended its original definitions in various directions. Divergent thinking is encouraged in creative practices. Methods of promoting divergent thinking and correlations of divergent thinking with other cognitive styles and factors still are well-represented in current trending research topics (Brown, 2009; Coskun, 2005; Larsen, 2013).

- **Complexity-Simplicity or conceptual articulation or cognitive complexity or conceptual differentiation**  
  *The tendency to prefer complex or simple conception constructs. More similar or related construct elements are considered simpler.*  
  This cognitive style is defined by the individual differences in personal constructs people employ to interpret human behaviour. Two commonly used metric are Bieri (1955)’s repertory grid developed based on Kelly (1963)’s work, and Crockett (1965)’s Role Category Questionnaire (RCQ).

- **Compartmentalization-Integration or conceptual integration**  
  *The tendency to perceive different constructs as more integrated or compartmentalized concepts.*  
  This cognitive style is considered to be functionally independent from conceptual differentiation (Fauconnier, 1998; Landfield, 1980; Messick and Kogan, 1963; Neimeyer et al., 1983). The term “conceptual integration” is also used in cognitive science referring to the mental
operation of blending. Like framing and chunking, this cognitive operation is considered dynamic, routine but mostly unconscious.

- **Impulsivity-Reflectivity** or **conceptual tempo**  
  *The rate one completes mental tasks. Reflective individuals tend to spend more time in considering alternative solutions than impulsive ones.*
  
  This cognitive style is commonly tested with variations of the Stroop test and Matching Familiar Figures Test (MFFT) and is more related with basic perceptual processing. However, these metrics are thought to be applicable in reflecting higher-level cognitive processes such as decision making. Correlation studies of gambling and weight control behaviour show possible associations between perceptual conceptual tempo with high-level cognitive tasks (Delgado-Rico et al., 2012; Gozali, 1969; Kagan, 1966).

- **Abstract-Concrete** or **conceptual complexity** (Gregorc, 1982; Halford, 1992; Schroder, 1971) or **Differentiation-Abstraction** (Gardner and Schoen, 1962)  
  *The preference of perceiving information.*

- **Random-Sequential** (Gregorc, 1982)  
  *The preference of ordering information.*

- **Leveler-Sharpener** (Holzman and Klein, 1954)  
  *The behaviour of assimilate new information with prior knowledge in either blending or differentiating to past memory.*

- **Adaptation-Innovation** (Berman et al., 1999; Kirton, 1976, 1984)  
  *The tendency to either adapt to existing rules or innovate to create new rules.*

Other examples of cognitive styles are **Holists-Serialists** (Clewley et al., 2010; Ford, 2000), **Verbaliser-Imager** (Pektas and Pultar, 2006), **Tolerance for instability / unrealistic experience** (Klein et al., 1962), **External-Internal Locus of Control** (Rotter, 1966), **Range of scanning** (Cas-sidy, 2004; Gardner and Long, 1962), **Breadth of categorization** (Davis et al., 1992; Ryan and Deci, 2000), **Extrinsic-Intrinsic Motivation** (Vallerand et al., 1992), **Mobility-Fixity** (Kozhevnikov et al., 2014; Witkin and Goodenough, 1981).

**Grouping Cognitive Styles Meaningfully**

Not only the cognitive style definitions are confusing, the higher level classifications are also confusing and lack of agreement. For example, some researchers categorizes cognitive styles into cognition-centered and activity-centered ones (such as learning styles) (Sternberg and Grigorenko, 1997), while others exclude learning styles from cognitive styles (Riding and Cheema, 1991) for their difference in number of style elements considered. Traditionally, cognitive styles are bipolar dimensions (as listed in the above subsection), while learning styles can have multiple types that
are do not fall into the “either-or” descriptions of the cognitive style dimensions. We argue that this differentiation based on number of elements is unnecessary, and advocate to include all types of cognitive styles in future frameworks. Because even in bipolar dimensions studies, it is difficult to arbitrarily split the subjects into two extremes and sometimes subjects are placed into several bins as opposed to just two groups. We see bipolar styles as a subset of all styles that can have two or more elements.

Another example of classification ambiguity is styles and strategies. Some researchers use these terms interchangeably, while some differentiate the two concepts by the degree of stability in an individual over time. Riding and Cheema (1991) argue that cognitive styles are attributes of how an individual performs cognitive tasks, and cannot be changed easily; while strategies can be learned and trained to change. While this is a convenient differentiation, we think it is not a sound one. One of the less disputed cognitive style dimension definition is “convergent-divergent thinking”. While some holds that this style is a relatively fixed attribute for an individual, applied research in design and business decision-making are suggesting that these thinking styles can be trained. Without getting into the more complex and traditional nature-nurture debate, we are picking our safe ground on not differentiating cognitive styles and strategies conceptually. We believe that all cognitive styles or strategies more or less have both nature and nurture components to it, and are sensitive to environmental changes over time. e.g. For different tasks, convergent or divergent thinking can be applied from the same person.

Some researchers like Zhang (2008) make strict associations between different cognitive style dimensions in their classification. The intellectual style types (I, II, III) are unique combinations of specific values in each dimension. For example, Type I takes the values of holistic, intuitive, perceiving, concrete random, innovation, reflectivity, divergent thinking, field independent at the same time, while Type II takes the values of analytic, sensing, judging, concrete sequential, adaption, impulsivity, convergent thinking, field dependent (Zhang, 2008, Appendix B). The assumption of associations between different dimensions is debatable. The three types are very limited in modeling individuals with other combination values, for example someone that is both intuitive and impulsive fits none of the Type I, II or III. We argue for non-bounding cognitive styles dimension classification to avoid this problem.

In applied cognitive styles research, studies that attempt to identify correlations between different dimensions are also cyclically popular.

**Seeking a Coherent Cognitive Styles Framework**

Kozhevnikov et al. (2014) argue that most of previous frameworks attempting to classify exploding cognitive styles research are lacking a unifying theory for cognitive styles. The authors provide an overview of current trends in cognitive style research towards hierarchical multilevel models, and proposes an integrated matrix taxonomy for cognitive styles. The authors adapt Nosal (1990)’s information processing theory to classify all cognitive styles by four levels of processing: 1) perception, 2) concept formation, 3) high-order cognitive processing, and 4) meta-cognitive pro-
cessing. With these four traditional levels on the vertical axis of the matrix taxonomy, the authors established the horizontal axis to be a new classification for different families of cognitive styles: 1) **context dependence vs. independence**, 2) **rule-based vs. intuitive processing**, 3) **internal vs. external locus of processing**, and 4) **integration vs. compartmentalization**. This taxonomy is illuminating in providing a big picture to cognitive styles research, and suggests research in applied fields should expand beyond focusing on one particular cognitive style (or style family). In this thesis, we are basing our classifications of cognitive styles in design decision making on Kozhevnikov et al. (2014)’s work.

Some of the later design decision-making theories draws and expands this body work, further proves that (at least under certain experimental settings), regardless of personal style, the assigned method of decision-making yields significant different outcomes. It is evidence that certain strategies can be trained for specific outcomes. For example, in one study demonstrated showing partial photographs vs. full photographs of inspiring works can reduce design fixation during the design progress (Cheng et al., 2014). Putting into the context of cognitive style, that might mean using a compartmentalized frame of reference in design might be more beneficial than an integrated frame of reference when looking for inspirations.

So far, we have presented our set of assumptions with cognitive styles:

1. Cognitive styles are relatively stable characteristics of individuals' preferred ways of processing information
2. Although generally defined as a general catch-all concept, cognitive styles are actually sensitive to its environment and can express different modes depending on the task or conditions
3. There are correlations between cognitive styles specific to tasks and task environments
4. At least some cognitive styles can be influenced through training to achieve different task outcomes (e.g. better design performance when trained to use more divergent thinking in brainstorming sessions.)

There are still many unknowns with this argument or research approach:

- **If Assumption (2) above is true, then:**
  - The cross-domain design generalizations must be attested individually to establish any claims.
  - Cognitive styles relevant to design decision making may or may not be the same for generative design decision making when both the nature of the task and task environment are changed.

- **Differentiating correlation and causation of cognitive styles and task outcomes**
  - More studies are correlation studies (probably in a questionable low number of participants as well). These studies only establish a possible link between cognitive and the task outcomes with high probabilities of confounds being the real causation.
Only more intervention studies where specific cognitive styles of thought are assigned to different groups of participants may increase the confidence of concluding a stronger link between cognitive styles and task outcomes.

2.5.2 (Ir)rational Decision Making

While investigating an issue of design decision-making, we must look into the vast body of decision-making literature, in various fields. The field of behavioral decision research is the study of judgment and decision making encompassing three interrelated categories of research: 1) normative analysis, 2) descriptive studies and 3) prescriptive interventions.

Historical perspectives in utility theory are based on pure rational decisions are considered as normative analysis, were intended to explain decision-making behavior, but have been widely accepted as wrong and used to identify best course of action for decision makers. Descriptive studies are studies that examine actual human behaviours, and normally are compared against normative methods for discrepancies and explanations. Prescriptive interventions attempt to alter people’s decision making behaviors by introducing factors that will bridge the gap between ideal rational models and the actual behavior models.

Progressions from utility theory to the bounded reality model (Herbert Simon) then to Prospect Theory (Kahneman and Tversky) s marking the clear realization of how the human mind is not strictly rational (Kahneman, 2012; Simon, 1996). Recent research in behavior research, especially those driven in economics / consumer behaviour, are providing more evidences to our “irrational” side in decision making. Ariely (2008)’s book Predictably Irrational is a collection of work that reflects research done in this area. Sheena Iyengar’s work on choice also reflects factors that can’t be explained by rationality. For example, when given more choices, people are less likely to make purchase decision. The term naturalistic decision making captures the movement away from the rational decision making paradigm.

Other formative or descriptive research have investigated into different decision making strategies and correlations between the strategies and decision outcomes, such as confidence in decision, rationality in decision etc. A comprehensive taxonomy by Beach and Mitchell (1978) is as following:

- **Aided-Analytic.** A process involves the use of tools (e.g. pencil, paper, computer programs, other interfaces) Normally involves listing pros and cons of alternatives, decomposing problem through use of decision trees etc. e.g. Benjamin Franklin’s moral algebra.

- **Unaided-Analytic.** Attempt to explore dimensions of the problem in the mind with no tools available. Most variants of the Subject Expected Utility (SEU) strategies that lives entirely in the decision-maker’s mind are considered belonging to this category.
  - **Compensatory.** Less of one component affecting the decision can be compensated by more in other components. The most famous of this category is Tversky’s additive
difference model. It explains intransitivity between decision. In other words, even if A is preferred over B, and B over C, still A is not necessarily preferred over C. Tversky’s compensatory model considers component-wise differences over all components.

– Noncompensatory. The most famous out of this category is Simon’s satisficing strategy. It selects the first decision alternative that exceeds the minimum bar of all components. It does not seek for the maximization, rather sufficiency. There are no trade-offs between components like compensatory strategies. Less complex versions of noncompensatory strategies include elimination by aspects and lexicographic. The elimination method is to select by one aspect and eliminates ones that doesn’t fit. The lexicographic method is to select the most important aspect and eliminates all alternatives except the superior one; if non is superior than moves to the next aspect. This method reduces information processing significantly.

– Mental movies / “scripts”. This method involves the decision maker imagining how things might be if this or that decision alternative were chosen and then picks the alternative for which the script turns out the best.

* Non-Analytic. This category involves simple, preformulated rules that are applied by rote to decision tasks. Examples include random selection (e.g. flipping a coin or rolling a dice, “I’m feeling lucky” while googling), by experience, intuition, convention/habit.

This 1978 classification of decision-making taxonomy is useful, however, a little bit out of date. For example, the aided and unaided differentiation is now blurred with the popularization of computer-aided tools. Many problems are evolving to be too complex to be managed “unaided” and are growing to require more than one decision maker to make decisions together. The boundaries of aided and unaided are still important in the investigation of decision making, especially in questions concerns human cognitive limits (e.g. how many is too many alternatives), but should not be ruling out listed “unaided” strategies to be used in an “aided environment”. For example, many decision making support systems are incorporating SEU or lexicographic capacities to assist complex institutional and collaborative decision making.

### 2.5.3 Domain of Design

Many research has been done to compare engineers to architects, scientists to artists, to show differences in the way they solve design problems. There are mixed results for whether the domain of design is an variable that affects the design process and outcomes. For example, Akin (2001) claims that architects breadth-first-then-depth behavior only exhibit in engineers, but not in architects. While more findings show that designers across different domains have shared commonalities, such as architects, product designers, software and mechanical engineering (Reymen et al., 2006; Thomas and Carroll, 1979; Visser, 2009), and the differences are only in terminology. Gross (2003) states the exact commonalities between designing a piece of software and a building.
In this thesis, we will continue to argue that there at least are some high-level abstractions are shared across all design activities in various domains. And we aim to explore through our small section of design studies to see which might be the common denominator and which might not be.

2.5.4 Expertise in Design

Years of design research have focused on how expert intuition is developed, and what are the differences between novice and expert design processes, how to identify experts and how to acquire expertise. In design literature, it is common to draw to expertise models from other fields. Schön (1983), Cross (2004), Lawson and Dorst (2009) commonly draw references from musicians, artists and chess players to illustrate their model on expertise.

Differences Between Novices and Experts

Numerous protocol studies and case studies have been conducted to investigate into the differences between novices and experts. Some showing conflicting results. Some of the better known findings are:

- Novices use more ‘trail and error’ techniques, generate one and evaluate, fail, then generate more. Experts make preliminary evaluation of all alternatives, then implement and finally makes a decision (Ahmed et al., 2003).

- Novices are more likely to rate distant-domain inspirations higher. Experts are more likely to rate near-domain inspirations higher (Ozkan and Dogan, 2013). Opposite claims are also found (Casakin, 2004).

- Experts are more likely to add constraints to the design than novices (Casakin, 2004).

- Experts use working-forward strategies. Novices use working-backwards strategies (Anderson, 2004). Opposite claims are also found (Ho and Group, 2001).

- Experts tend to use more breadth-first, depth-next approach. Novices use mainly depth searches in seeking alternatives (Woodbury and Burrow, 2006b). Claims of experts employing more depth-first approaches can achieve same qualities with those use breadth-first (Cross, 2004).

- Experts tend to group problems based on deep structures and organizing problems with the same underlying principles together, while novices focus on surface features (Chi et al., 1979).

- Experts are more cognitively active and productive compared to novices (Kavakli and Gero, 2001).

- Experts are more likely to question data and engage in more reflections (Ahmed et al., 2003; Schön, 1983).
• Experts are more likely to employ problem decomposition, and do so with explicit strategies (Ball et al., 2004).

Expert Intuition

As discussed in 2.1, the design process is part empirical and part tacit. Design expertise also correlates with both empirical and tacit knowledge acquisition. Empirical or formal knowledge can be obtained through formal training, books or other explicit forms of instruction. Tacit knowledge, on the other hand, is less understood and more difficult to obtain. Eraut defines tacit knowledge in three forms (Eraut, 2000): 1) situational understanding, 2) standard, routinised procedures and 3) increasingly intuitive decision-making. How do expert acquire tacit knowledge while they grow from novices is still not very well understood.

An interesting debate on expert intuition is between Kahneman and Klein. Kahneman believes that expert intuitions are overrated, and not reliable. He found that financial adviser’s performance from year-to-year has a 0.01 correlation, which means there are no correlation between experience and investment outcomes. This means performance is not based on skill, but rather sheer luck from year-to-year. He further found that most companies reward this chance thinking it’s skill, reinforces an illusion of skill that causes overconfidence in financial experts.

“For a large majority of fund managers, the selection of stocks is more like rolling dice than like playing poker. Typically at least two out of every three mutual funds underperform the overall market in any given year.” (Kahneman, 2012, page 215).

On the other side of the table, Klein (2013) represents the other school that believes that expert intuitions are not only accurate, but are faster than rational judgments in many circumstances. His studies with fire fighters have shown that experience experts have the almost instinct reaction to get out of a danger zone split seconds before the building collapses, without conscious realization of how they came to the conclusion.

The two seemingly contradicting schools, Kahneman and Klein (2009), eventually came into agreement and wrote a paper together on their “failure to disagree” with each other. In which they concluded that expert intuition can be developed for events that have recurring characteristics such as fighting reoccurring fire conditions, but not for those events that are random such as the market with a fluctuating pattern different year-to-year. This conclusion seems quite intuitive at a glance, but how to identify which of the events or tasks are repeatable that develops expert intuition still remains difficult. Afterall, managing mutual funds was considered to be a skills trade. In this thesis, we argue that design as an activity shared across disciplines encompass both tasks that are repeatable and those are not. One of the key goals of design process studies to us is to identify which processes fall under Kahneman’s model and which ones under Klein’s.
Measuring Design Expertise

Dreyfus’ Model of Skill Acquisition is a recognized model adopted for discussions in expertise across many domains (Benner, 2004). In the domain of design, it is designated to identify the different cognitive behaviours and strategies used by designers at various levels of expertise. In this thesis, we are adopting this model in Study 1 (Chapter 3) for self-evaluation of expertise Table 2.1.

Table 2.1: Description of Dreyfus’ Model of Skill Acquisition (Five Stage Cognitive Model) (Cheetham and Chivers, 2005)

<table>
<thead>
<tr>
<th>Expertise Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novice</td>
<td>“rigid adherence to taught rules or plans”; no exercise of “discretionary judgment”</td>
</tr>
<tr>
<td>Advanced Beginner</td>
<td>limited “situational perception”; all aspects of work treated separately with equal importance</td>
</tr>
<tr>
<td>Competent</td>
<td>“coping with crowdedness” (multiple activities, accumulation of information); some perception of actions in relation to goals; deliberate planning; formulates routines</td>
</tr>
<tr>
<td>Proficient</td>
<td>holistic view of situation; prioritizes importance of aspects; “perceives deviations from the normal pattern”; employs maxims for guidance, with meanings that adapt to the situation at hand</td>
</tr>
<tr>
<td>Expert</td>
<td>transcends reliance on rules, guidelines, and maxims; “intuitive grasp of situations based on deep, tacit understanding”; has “vision of what is possible”; uses “analytical approaches” in new situations or in case of problems</td>
</tr>
</tbody>
</table>

2.5.5 Other Possible Confounding Variables: Not In Scope

The above variables discussed are the main focus of this thesis, but are not the only variables that can affect designers’ process and outcomes. It is worth mentioning other possible confounding variables that are not in scope of this thesis. Researchers have reported correlations of design performance with variables such as gender (Allinson and Hayes, 1996), culture (Ji et al., 2000; Miyamoto et al., 2006). In our studies, we do not recruit for or against any of these variables, and hoping the random sampling to be naturally representative of the studied population and balances itself out.

2.5.6 Applied Visual Analytics

Visual Analytics (VA) is “the science of analytical reasoning facilitated by interactive visual interfaces” (Thomas and Cook, 2005). This multidisciplinary field associates with analyzing large-scaled data and present with interactive visual representations to assist with human cognitive limits for interpreting the data. Utilizing VA techniques in generative design computational tools may be very beneficial, especially in assisting designers to explore large number of alternatives. Interface recommendations based on findings from this thesis attempt to incorporate VA for supporting designers to make better design decisions.
Aside from using VA in computational tools, this thesis also includes an interactive visualization built for the purpose of data analysis (Study 3, Chapter 5. It helps the researcher to analyze video codes from observational design study to understand user design decision making. Its static visual forms also assisted in presenting and explaining the analysis results in this thesis.

2.6 Summary

This chapter summarized related literature for the following studies in design, design exploration, and design alternatives. It also touched on common variables in design studies that will be explored in the following studies, such as cognitive styles, decision making and design expertise. In following chapters that describe each specific study, I will omit details on literature review and refer to relevant sections in this chapter when related works are needed to back up my viewpoints.
Chapter 3

Study 1: Design Exploration With Website Conceptual Sketching

3.1 Introduction

An ability to generate alternatives is commonly accepted as a necessary and sufficient condition for design expertise. Approaches to design exploration, especially how alternatives are generated and evaluated during the process, have been studied by researchers across various disciplines (Section 2.4). Many studies investigated differences in design processes between expert and novice designers, with the intention of identifying more expert-like design behaviours. Similar to these earlier studies, we initially planned to investigate measurable differences between novice and experienced designers in this exploratory study. Eighteen participants performed a website conceptual design task via sketching. Qualitative and quantitative analysis revealed richer differences than typical bipolar scale differentiations, e.g. breadth-first vs. depth-first, problem-based vs. solution-based. Our findings suggest a need to revisit previous studies by re-interpreting the results with a broader investigation on design processes in general. This chapter presents observations of different styles of alternative generation, and their relations with other factors, including novelty and usability).

3.1.1 Research Questions

Based on existing literature (discussed in Chapter 2 Related Works section) and findings from the original investigation of measurable differences between novice and experienced designers, this study reframed its research questions and hypotheses to be revisited using both quantitative and qualitative research methods. This new position is more diverse than just focusing on differentiating bipolar categories such as novice/expert. It intends to investigate possible variances in different design processes (e.g. style of alternative generation, number of alternatives generated etc.), and design background (e.g. occupation, experience measured through different self-report measures), with regard to design outcomes (e.g. both self and expert panel rated novelty and usability, confidence level, etc.).
RQ1 How do designers differ in terms of their processes of generating alternatives?

RQ2 How does the number of alternatives generated during the design process relate with design outcomes and designer background?

RQ3 Are there any correlations between identified design processes and designer background with design outcomes?

RQ4 How do designer’s self-perception in expertise-level confidence relate to his or her actual experience in task domain and task performance?

3.1.2 Quantitative Research Hypotheses

H1 More experienced designers generate a larger number of alternatives.

H2 More experienced designers are more likely to generate alternatives in parallel, rather than sequentially, than novice designers.

H3 Number of alternatives correlate positively with design outcomes (i.e. novelty and usability) regardless of experience.

H4 Designers that generate more alternatives have more confidence about their task performance.

H5 Both more experienced and less experienced designers will be inaccurate in assessing their own performance level.

3.2 Methods

3.2.1 Study Setup

In this study, participants used a pen and pieces of letter-sized printing paper to produce sketches for an online shopping website. A Flip UltraHD Video Camera was set up to record the participants’ sketching actions and the post-task presentation and interview. The video camera was set up to record at an angle so that the participant’s face was not in the picture.

3.2.2 Participants

Participants were recruited separately from Craigslist, the software industry, and among graduate students at Simon Fraser University. No pre-screening of the participants was done, as the recruitment notice intentionally did not mention the true study objective relating to design expertise. Best efforts were made to avoid biasing participant behaviour by revealing the study objectives (e.g., participants may impose self-impressions of novice or expert that can affect their task behaviour, or participants may try to impress others unconsciously by using what is thought to be “expert-like” strategies).
Among the 18 participants, there were 8 female, 10 male; 8 had previous experience in designing website and 10 had no experience. The mean age of the sample was 33.64 (SD=14.45). All participants were frequent computer users. Average computer use per week ranged from 14 to 130 hours ($M=52.8$, $SD=34.29$). Average Internet browsing time per week ranged from 5 to 80 ($M=27.18$, $SD=20.16$). Note that reports of computer use per week and internet browsing estimates appeared to be unrealistically high for some.

3.2.3 Procedure

Each participant was first introduced to study. Informed consent for the study, as well as permission to video record, were obtained. Participants were asked to design a website for online shopping for a local department store like Sears or The Bay in 30 minutes. Participants were instructed to use the full 30 minutes for the task, even if they finished designing earlier. A timer was used to time the participants. Participants were free to make any assumptions they wanted to about the site, but were asked to note them and explain at the end of the design sessions. They were told to use as many or as little scratch pieces of paper as they wished. No instructions with wording related to design alternatives, usability, or novelty were given to avoid biases.

After the sketching task, participants were asked to spend five minutes to present their design(s) to the researcher. Participants explained their design concepts and the details of each sketch produced. Then a post-task questionnaire containing background questions, as well as the above self-reported measures, was given to the participant. Lastly, a short informal interview was conducted to probe further into the answers from the questionnaires (e.g. the detailed design alternative generation sequence).

All user sketches were collected, scanned, and labeled with participant numbers. Comments about design intent by participants were transcribed from the post-task design presentation and provided to the expert-panel along with the sketches. This was done to avoid biases in listening to the audio of the presentations, as impressions on the participant’s age and quality of speech can affect expert judgment. The expert panel then rated the sketches in terms of novelty, usability, and vagueness as described in section 3.2.4.

3.2.4 Expert Panel

Panel Members

An independent expert panel with two experts were called to judge the design outcomes of the participants. Both experts were user experience designers at a major IT company, and mainly worked on web products. One had 10 years of experience in the domain, while the other had 4 years. Both were considered domain experts for the interface design task in this study, and were expected to be able to judge the quality of the designs generated by the participants.


Judging Process

The researchers explicitly asked the expert panels to rate designs with three scores: vagueness, novelty and usability, but left it to expert opinions to decide detailed judging criteria for each scores. To avoid bias, the expert panel first met to discuss detailed judging criteria before seeing the sketches. Then each of them rated all the sketches independently (i.e. with no communications and in separate rooms without seeing each other) using the agreed set of judging criteria. After that, they met again to discuss each sketch, and gave a joint score based on their discussions. This joint score was deemed to be free of inter-rater biases.

The experts had no background information about the participants (e.g. occupation, experience level etc.) and were only showed sketches. They rated all sketches on usability, novelty, and vagueness, on a scale of 0 to 10.

Criteria For Judging

Usability  The expert panel chose to use a book on usability (Krug, 2006) and an internal corporate usability guideline matrix used at the experts’ company as the basis for discussing a set of agreeable criteria for usability. The final set of criteria contained the following list:

- Clear visual hierarchy on each page
- Users have clear choices
- What users are looking for is obvious
- Users know where to start
- Visibility of system status
- Match between system and the real world
- User control and freedom
- Consistency and standards
- Error prevention
- Recognition rather than recall
- Flexibility and efficiency of use
- Help and documentation

Novelty  The experts relied on their own domain knowledge of what is conventional and what was new. In this case it was difficult to outline detailed criteria for it, as it was tacit knowledge based on each expert’s judgment.

Vagueness  The sketches were rated by the expert according to vagueness. This is defined as the amount of details to be added before the design could be implemented as an actual web page.
3.2.5 Experimental Variables

All variables in this observational study were measured or reported, as opposed to assigned or manipulated. Predictor variables and response variables were selected to reflect the research question.

Predictor Variables

Perceived style of alternative generation  This was a reported measure from each participant’s post-task questionnaire (Appendix A). It probed how the participants generated design alternatives. There are four values to this nominal variable: a) no alternatives, the participant only conceptualized and sketched one design, b) reported to have more than one alternatives, but only sketched one, c) sketched multiple alternatives in parallel, and d) sketched multiple alternatives sequentially. Participants were asked to report the total number of alternatives sketched on paper and the number of alternatives thought about but did not sketch.

Initial task action  This was a self-reported response on the post-task questionnaires. This response had five options to choose from that described the initial reaction after the subject first received the task (Listed in 3.2).

Occupation  This is an open-ended background question on the post-task questionnaire.

Experience in task domain  Participant’s post-task questionnaire contained a yes/no question inquiring about past experience in designing websites. If yes was answered, a rough estimation on the number of projects and hours spent was requested. In hind-sight the number estimations are too inaccurate to be used as a continuous numeric datum so it was converted it into an ordinal variable. The participants were classified into three groups: 1) no experience, 2) only unpaid project experiences, and 3) both unpaid and paid professional experiences.

Response Variables

Self-reported measures were all based on a 7-point Likert scale following psychological test standards, except for self-judged expertise level, which was based on the Dreyfus Model (Table 2.1). Expert-rated scores were all based on 10-point scales, following conventional marking practices.

Self-reported difficulty of the task  (Q2) This was a self-reported measure from the post-task questionnaire, based on a 7-point Likert scale. It reflected the task difficulty perceived by the participant with 1 being very easy and 7 being very difficult.

Self-reported confidence level  (Q4) This was a self-reported confidence rating from the post-task questionnaire, based on a 7-point Likert scale. It reflected how confident the participant felt about their final design, with 1 being very confident and 7 being very unconfident.
**Self-reported usability**  (Q6) This was a self-reported usability rating from the post-task questionnaire, also based on a 7-point Likert scale. The participant is asked to rate the ease of use of the final design, with 1 being very low usability and 7 being very high usability.

**Self-reported novelty**  (Q7) This was a self-reported novelty rating from the post-task questionnaire, also based on a 7-point Likert scale with 1 being very conventional and 7 being very novel.

**Self-judged expertise level**  (Q5) This was a self-reported level of expertise based on the Dreyfus’ Model of Skill Acquisition (shown in 2.1, discussed in detail in Section 2.5.4) provided to the participants. It contained a scale of 1 to 9, with 1 being novice and 9 being expert. In-between values represent different expertise levels.

**Expert-rated usability**  Usability score judged by the expert panel for the sketched designs from each participant, ranging from 0 to 10. This was based on a pre-agreed set of criteria.

**Expert-rated novelty**  Novelty score judged by the expert panel for the sketched designs from each participant, ranging from 0 to 10.

**Expert-rated vagueness**  This is a scale of 0 to 10, rated by the expert on the general vagueness of the sketch.

### 3.3 Results

#### 3.3.1 Expert Panel Ratings

Inter-rater correlations were calculated between the two experts on the panel for their independent ratings for usability, novelty, and vagueness of the sketches (Figure 3.1, Left). No satisfactory inter-rater consistency was found for any of the three scores. Therefore, the experts conducted a meeting to discuss the scores and came to a consensus with a joint panel score. In the following section, any expert-rated scores refer to the joint score given by the panel agreement. A distribution of the joint scores are presented in Figure 3.2. However, it is interesting to note the consensus score had a very high correlation with the averaged score between the two (Figure 3.1, Right).

#### 3.3.2 Self-Reported vs. Expert-Rated Scores

In an attempt to identify a reliable and easily measurable difference between novice and expert designers, the self-reported and expert-rated scores were compared for correlation. No significant correlation was found in usability, as shown in Figure 3.3. However, a significant correlation was found for novelty. Hence, self-reported usability can not be used as a reliable measure to predict design outcomes, while self-reported novelty might be.
Figure 3.1: **Left:** Inter-rater correlations between two experts. **Vagueness:** Correlation = 0.52, \( p = 0.027^* \). **Novelty:** Correlation = 0.67, \( p = 0.002^* \). **Usability:** Correlation = 0.48, \( p = 0.043^* \). **Right:** Correlations between joint scores produced via expert consensus or simply averaging the two scores. Consensus score was produced during a meeting after both experts individually rated the sketches. Discussions during the meeting rendered a final score that both experts agreed on. **Vagueness:** Correlation = 0.95, \( p < 0.001^* \). **Novelty:** Correlation = 0.95, \( p < 0.001^* \). **Usability:** Correlation = 0.99, \( p < 0.001^* \).

![Correlation Table]

<table>
<thead>
<tr>
<th></th>
<th>( r )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Vagueness</strong></td>
<td>0.52</td>
<td>0.027*</td>
</tr>
<tr>
<td><strong>Novelty</strong></td>
<td>0.67</td>
<td>0.002*</td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td>0.48</td>
<td>0.043*</td>
</tr>
<tr>
<td><strong>Consensus</strong></td>
<td>0.95</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>0.99</td>
<td>&lt;0.001*</td>
</tr>
</tbody>
</table>

Figure 3.2: Distributions of joint scores through expert panel consensus for vagueness, novelty and usability (\( n = 18 \)). Error bars showing standard error.

![Distribution Graphs]

Figure 3.3: Correlations between self-rated and expert-rated scores for novelty and usability. **Novelty:** Correlation = 0.65, \( p = 0.009^* \). **Usability:** Correlation = 0.40, \( p = 0.500 \).

![Correlation Table]

<table>
<thead>
<tr>
<th></th>
<th>( r )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Novelty</strong></td>
<td>0.65</td>
<td>0.009*</td>
</tr>
<tr>
<td><strong>Usability</strong></td>
<td>0.40</td>
<td>0.500</td>
</tr>
</tbody>
</table>

42
3.3.3 Effects of Predictor Variables

Data are analyzed using a factorial MANOVA with four between-groups factors (style of design alternative generation, initial task action, occupation, and experience in task domain). The analysis revealed no significant interactions between the four factors in any factorial combinations.

A subsequent MANOVA analysis was performed on each individual factor for the effect of the following expert rated response variables: vagueness, novelty, usability, and self-reported response variables: usability, novelty, confidence level, expertise level, perceived task difficulty. Details for each predictor variable are discussed in detail in the next subsection.

Significant Effect: Style of Design Alternative Generation

An individual MANOVA revealed a significant multivariate effect for the style of design alternative generation, Wilks’ $\lambda = 0.009, F(24, 21) = 3.60, p = 0.002^*$, on the above-mentioned response variables. A univariate ANOVA of each response variable was performed (as discussed below). Tukey’s HSD multiple comparisons post-hoc test was performed for each response variable that showed statistically significance. Response variables that showed a significant effect are displayed in expert-rated vagueness, expert-rated novelty, and self-reported usability. The analysis revealed no statistical difference among groups of different styles of alternative generation (described in section 3.2.5) for expert-rated usability, self-perceived novelty, confidence-level, expertise level, and self-perceived difficulty of the task. Table 3.1 shows a summary of all the means of related scores classified by style of design alternative generation. Note that No Alternatives means there was only one solution.

Table 3.1: Means of Responses. (E): expert-rated. (S): self-reported.

<table>
<thead>
<tr>
<th>Style of Alt. Generation</th>
<th>#of ppl</th>
<th>Vag. (E)</th>
<th>Nov. (E)</th>
<th>Usab. (E)</th>
<th>Nov. (S)</th>
<th>Usab. (S)</th>
<th>Conf. (S)</th>
<th>Exp. (S)</th>
<th>Diff. (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) No Alternatives</td>
<td>11</td>
<td>3.68</td>
<td>3.91</td>
<td>4.82</td>
<td>2.91</td>
<td>2.91</td>
<td>5.36</td>
<td>4.72</td>
<td>2.91</td>
</tr>
<tr>
<td>b) Multi: 1 Sketched</td>
<td>2</td>
<td>5.75</td>
<td>3.75</td>
<td>7.00</td>
<td>3.00</td>
<td>1.50</td>
<td>4.50</td>
<td>5.50</td>
<td>2.50</td>
</tr>
<tr>
<td>c) Multi: Parallel</td>
<td>2</td>
<td>8.75</td>
<td>4.00</td>
<td>3.75</td>
<td>3.00</td>
<td>4.50</td>
<td>3.00</td>
<td>5.50</td>
<td>2.50</td>
</tr>
<tr>
<td>d) Multi: Sequential</td>
<td>3</td>
<td>6.50</td>
<td>8.67</td>
<td>6.50</td>
<td>5.33</td>
<td>2.67</td>
<td>6.67</td>
<td>8.00</td>
<td>3.33</td>
</tr>
</tbody>
</table>

Expert-rated Vagueness Participants who self-reported as working on multiple ideas in parallel (c), $M = 9.0, SE = 1.12$, had a significantly higher vagueness score than those with no alternatives (a), $M = 3.68, SE = 0.47$. No significant differences were found between other groups (Figure 3.4).
**Expert-rated Novelty**  Participants self-reported as working on multiple ideas sequentially (d), $M = 8.66, SE = 1.14$, had a significantly higher novelty score than those with no alternatives (a), $M = 3.91, SE = 0.60$. No significant differences were found between other groups (Figure 3.5).

**Self-reported Usability**  Participants who self-reported as working on multiple ideas sequentially (d), $(M = 6.67, SE = 0.61)$, reported significantly higher on self-perceived usability than those who in parallel (c), $M = 3.00, SE = 0.75$. It is interesting to note the participants that worked on multiple alternatives (c) in parallel had the lowest self-perceived usability (Figure 3.6).
No Alternatives vs. Multiple Alternatives (Aggregated Group Comparison)  The above analysis may have been affected by low numbers of participants, and uneven sampling (n is not equal) among observed groups in (b), (c) and (d), as shown Table 3.1. A more powerful statistical test (t-test) was performed by combining (a) + (b) into one category by self-reported style of alternative generation: no alternatives produced on paper (however, group (b) reported having multiple alternatives in their mind). Likewise, (c) + (d) are combined into one category: multiple alternatives produced (despite different reported generation processes).

The two groups (no alternatives vs. multiple alternatives) were compared for all response variables. Significant differences were found for expert-rated vagueness (Figure 3.7) and expert-rated novelty (Figure 3.8) scores. No significant differences were found for the remaining response variables: expert-rated usability, self-rated novelty, usability, expertise level, confidence level, and perceived task difficulty.

![Figure 3.7: Significant difference in expert-rated vagueness scores for no alternatives (M = 4.00, SD = 1.81, n = 13) and multiple alternatives (M = 7.40, SD = 1.43, n = 5), t(16) = 3.74, p = 0.002**)](image1)

![Figure 3.8: Significant difference in expert-rated novelty scores for no alternatives (M = 3.88, SD = 1.75, n = 13) and multiple alternatives (M = 6.80, SD = 3.35, n = 5, t(16) = 2.45, p = 0.026*)](image2)
With no aggregated groups, the total number of alternatives generated was treated as a continuous variable to conduct fit model analysis. No linear fit relationships were found between the number of alternatives generated and any of the response variables, except for vagueness (discussed in Section 3.3.4 of Figure 3.11).

**Effect: Experience in Task Domain**

An individual MANOVA revealed a significant multivariate effect for the style of design alternative generation, Wilks’ $\lambda = 0.052$, $F(16, 16) = 3.39$, $p = 0.010^*$. A univariate ANOVA revealed only a significant difference between groups for self-reported expertise level based on the Dreyfus Model of Expertise (Figure 3.9, discussed in Related Works Section 2.5.4). No significant differences were found between groups for all other variables (i.e. novelty, vagueness, usability, confidence level, task difficulty).

![Figure 3.9: Means of Self-Judged Expertise scores (Y Axis) vs. Self-Reported Experience in Task Domain (X Axis). Significant main effect, $F(2, 17) = 22.60$, $p < 0.0001$* (a) no experience, $n = 9$, $M = 3.67$, $SE = 0.44$. (b) only unpaid experience, $n = 5$, $M = 6.80$, $SE = 0.49$. (c) with paid experience, $n = 4$, $M = 7.75$, $SE = 0.25$](image)

**No Effect: Initial Task Action**

Responses for initial task action were collected through the post-task questionnaire. No significant multivariate effect was found for initial task action, Wilks’ $\lambda = 0.062$, $F(28, 27) = 1.10$, $p = 0.40$. Hence, no further univariate ANOVA was performed for each response variable. The distribution of subject responses to initial task action are shown in Table 3.2.

**No Effect: Occupation**

Subjects provided open-ended responses to the occupation question. Responses are categorized into three groups: 1) designers, 2) science or engineering professionals, 3) others. No significant multivariate effect of occupation was found for all response variables, Wilks’ $\lambda = 0.224$, $F(16, 16) = 1.11$, $p = 0.42$. No further univariate ANOVA was performed for each response variable. However,
Table 3.2: Distribution of Initial Task Action Responses. Table showing total number of subjects for each initial task action response, as well as number of subjects categorized by style of alternative generation. Style of alternatives generation a) No Alternatives, b) Multi: 1 Sketched, c) Multi: Parallel, d) Multi: Sequential.

<table>
<thead>
<tr>
<th>Initial Task Action Response</th>
<th># of Subjects (Sum of N)</th>
<th>Style of Alt Gen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) felt lost, and did not know where to start</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2) recalled about similar websites I had seen / worked before, and followed their patterns</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>3) immediately followed a routine design pattern that I normally follow</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4) thought about what is the most important aspects to be addressed</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>5) followed my intuition</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

It is interesting to note the LS Means plot reveals a trend regarding expert-rated vagueness, where designers are producing significantly more vague sketches than science and engineering disciplines, $t = 2.13, p = 0.024^*$ (Figure 3.10).

Figure 3.10: Means of Expert-Rated Vagueness scores (Y Axis) vs. Occupation (X Axis). designers: $n = 7, M = 6.21, SE = 0.92$. others: $n = 7, M = 4.79, SE = 0.64$. science / engineering disciplines: $n = 4, M = 3.00, SE = 0.89$.

### 3.3.4 Other Speculations

Outside of the predictor versus response variable analysis, some variables that showed significant correlations with each other (Figure 3.11) through matched pair analysis:
Figure 3.11: Variables that showed significant correlations with each other.

- **Expert-rated usability and expert-rated novelty.** This correlation may be somewhat contradictory to current web usability design standards at first. Novel approaches are commonly avoided in web design to conform to existing user habits (Krug, 2006). Innovations can only become new conventions when enough users become used to them. This is now done through forced exposure through the tech giants’ associated products and services (e.g. Apple, Facebook, Google). The risk of high user resentment initially is usually expected. Even with forced exposure, many novel features do not become established over time, and have to be taken off the shelf if user complaints do not phase out with the learning curve. In a post-analysis follow-up with the expert panel on the possibility of this counter-intuitive correlation, the explanation reveals a difference in definitions of novelty. The researcher understood novel as being unconventional, which can include both good and bad designs, as based on previous design literature. However, the expert panel had reached a consensus that novel has a positive connotation, and should only include good unconventional designs. This suggests that for future studies along this line, more explicit defined criteria must be established, even if previous design literature commonly leaves it open for expert panels’ implicit judgment.

- **Number of total alternatives sketched and expert-rated vagueness score.** The likely explanation for this correlation is that within the constrained time frame, the more alternatives one sketches, the more vague the sketches. Traditionally, expert designers are associated with generating more alternatives that are also more vague. It is believed that the vagueness of the sketches reduces premature commitment to specific ideas, thereby allowing for more flexibility in design development, which eventually leads to more creative final solutions. A bold speculation is that vagueness may be a by-product of generating more alternatives due to time constraints and more matured initial concept generation style. Hence, generating more alternatives may or may not be the cause of more creative designs, but the vagueness generally caused by having more alternatives might be the root cause. In other words, an interesting question for future studies to ask is: in the conceptual design phase, does producing one extremely vague sketch gives more creative final solutions than producing a big number of concrete sketches?

- **Self-reported expertise rating and self-reported computer use.** Intuitively, this can be interpreted as subjects who surf the web more can be better at designing websites due to
exposure. On the flip side, it might also be an indication of false self-perception from the subjects - mistaking experience surfing websites with proficiency of designing websites. We believe both interpretations are reflective of this study.

Another interesting observation has probable implications towards the generalization of designer behaviour. The group of designers was observed to generate more alternatives than the other occupation groups regardless of their experience level in the task domain (Figure 3.12). The design group spans over multiple disciplines of design, including architecture, web, graphic and industrial designers. In this study, results imply that designers as a group share some characteristics in terms of style of alternative generations. Though we need more quantitative studies to confirm this hypothesis, we believe that this could be evidence for being “designerly” (Cross, 2001).

Figure 3.12: Heatmap of distribution of subject style of alternative generation by occupation and experience levels described in Section 3.2.5.

3.4 Discussion

RQ1. How do designers differ in terms of their processes of generating alternatives? Consistent with the literature, expert designers do generate more design alternatives than novices in this study. It is interesting to note that all novices in this study only generated one solution (H1 confirmed). However, the style of alternative generation differs among the expert designers. (Not enough evidence to confirm or reject H2. More styles of alternative generation are discovered).

H1: Observations consistent with H1 that more experienced designers generate a larger number of alternatives.

H2: There is not enough evidence to confirm or reject the hypothesis that more experienced designers are more likely to generate alternatives in parallel, rather than sequentially, than novice designers.

In Figure 3.13, we have listed the observed styles of alternative generation:

1. depth-only (no alternatives exploration), Figure 3.13 (1)
2. breadth-only (parallel exploration), Figure 3.13 (2)
3. breadth-the-depth (branching exploration, B2D), Figure 3.13 (3)
4. depth-then-breadth (sequential exploration, D2B), Figure 3.13 (4)

1) **Depth-Only** (No Exploration)

The participant only generates one solution, makes revisions to the single solution until he/she satisfies with it, or until time runs out.

No alternatives exploration is present in this case.

2) **Breadth-Only** (Parallel Exploration)

The participant tends to generate multiple alternatives from the beginning of the design process, then incrementally works on each of the alternative in parallel. No alternatives were rejected until final decision time.

Alternatives in parallel.

3) **Breadth-Then-Depth** (Branching Exploration)

The participant first generates multiple alternatives, then works on a selected one or more alternatives to continue exploration, and abandons the rest. Then based on the selected alternative(s), multiple new alternatives are generated as branching revisions to its parent alternative in various forms. The selection and generation cycle continues to repeat until a satisfying solution is found.

This appears to be a breadth-first-then-depth alternative exploration process. We name this as branching exploration, as it branches like a tree.

4) **Depth-Then-Breadth** (Sequential Exploration)

The participant first generates one alternative, then works on revising this alternative until he/she hits a wall, realizing this solution does not work, or can't be developed further. Then, the subject generates another alternative, continues to revise as the first alternative until he/she hits a wall again, and generates a new alternative. This process repeats until a satisfying solution is found.

This appears to be a depth-first-then-breadth alternative exploration process. We name this as sequential exploration, as it is a single-threaded process of one alternative at a time, as opposed to the parallel case.

Figure 3.13: Different alternative generation processes observed.
The alternative style naming is following references to depth-first or breadth-first search, or alternative generation patterns commonly seen in the design exploration and problem solving literature described in Chapter 2. Both breadth-only (parallel exploration) (Figure 3.13 (2)) (Lawson and Dorst, 2009) and breadth-first-then-depth (branching exploration), B2D (Figure 3.13 (3) (Woodbury and Burrow, 2006b) are known as expert design exploration behaviour. These two styles are observed only in two subjects, P14 and P18 respectively.

On the other hand, the D2B style, depth-then-breadth (sequential exploration) (Figure 3.13 (4)) has not been discussed by the literature, but was observed in 3 out of the 5 multiple alternatives generated cases. All D2B occurrences observed are from experienced designers with more than 5 years experience in the task domain. Example sketches from subjects using the D2B style are shown in Figure 3.14 and 3.15 (P1 and P13). We speculate based on our findings that the D2B style of alternative generation observed in experienced designers may be a new style of alternative generation that is worth further investigations.

Figure 3.14: Sample sketches from P1, showing 2 alternatives out of the total three produced. The visual tour alternative on the left aims to replicate the experience of shopping at a physical store location, by allowing the users to navigate through the merchandises by floors and brand display sections. The generic inventory alternative on the right has a conventional shopping site view.

Figure 3.15: Sample sketches from P13, showing all 4 alternatives produced. The subject extensively explored different site layouts and user interactions, such as animation flows, zooming behaviour, and drag-and-drop actions.
Similar design methodology studies can be conducted comparing the four styles of alternative generation techniques to see whether one promotes better design outcomes. Also, effects of different styles of alternative generation in different task environments (e.g. digital or physical prototype, with or without generative tools) need to be studied further to understand how to better support the design exploration process.

RQ2. How does the number of alternatives generated during the design process relate to design outcomes and designer background? No linear relationships were found between the number of alternatives generated and any of design outcomes or design background variables, except for vagueness. Figure 3.16 and Figure 3.17 shows sketches from two subjects that generated no alternatives in the task. These are evidently contain more details, such as details of text boxes, maps and information about shoes on sale. In contrast, Figure 3.18 shows sketches from a subject who generated 4 alternatives, and mostly concerned themselves with various interaction flows and layouts. No details of the sites were presented at all in the alternatives.

Aggregated group analysis of subjects that generated no alternatives vs. subjects that generated multiple alternatives revealed a significant difference in expert-rated novelty. (H3 confirmed for novelty and rejected for usability. H4 rejected.)

H3: Number of alternatives generated correlate positively with both expert-rated novelty, regardless of designer experience. In other words, if novice designers generated more alternatives, they can also achieve higher novelty scores. However, no correlations between number alternatives are found for usability.

H4: Observations show that designers who generate more alternatives are not more confidence about their task performance.

Figure 3.16: Sample sketches showing detailed sketching by a single solution participant (P6). It contains detailed components of the web page views.
Figure 3.17: Sample sketches showing detailed design by single solution participant (P10). It includes both the web page views and site structures.

Figure 3.18: Sample sketches showing four very vague and abstract alternatives by P18, exploring site layouts.
RQ3. Are there any correlations between identified design processes and background vs. design outcomes? From the sketches we are able to see that different participants went through drastically different design outcomes. Statistical analysis between style of alternative generation and design outcomes were performed. The most reliable measure was found to be novelty, which has a high level of agreement between two experts, and also correlates with subject self-reported score. This implies self-rated novelty scores may be useful as a measurement for self design evaluation.

The depth-then-breadth (sequential exploration) group produced significantly more novel alternative generation groups, and reported significantly higher self-perceived usability (but not in expert-rated usability). The breadth-then-depth approach (parallel exploration) was observed to have significantly lower expert-rated usability score, the lowest self-reported usability score, and significant higher expert-rated vagueness score. No statistical differences were found among different styles of alternative generations in other variables including experience in task domain, self-judged expertise level, confidence level, and difficulty of task.

It is also interesting to note the diversely different design processes, with other typically seen sketches of site layout and interaction flows, as well as user case scenario based approaches (Figure 3.19) and purely text-based designs (Figure 3.20).

![Figure 3.19: Sample sketches from P16 based on a user scenario approach.](image)

RQ4. How does a designer self-perception (i.e. expertise-level and confidence) relate to his or her actual experience in task domain and task performance? Results shown that the self-rated expertise scale based on Dreyfus’s model of expertise has strong correlations with self-reported project experience. This implies that designers, regardless of actual experience level, can relatively accurately identify themselves on a recognized scale.
However, the self-rated design outcomes scores (i.e. novelty and usability) have no evidence of being correlated with expert-rated design outcomes. This implies that designers, regardless of actual experience level, have a poor understanding or recognition of their own designs.

Self-reported confidence levels have not shown evidence of being related to design experience level, or design outcomes. This implies predicted design outcomes coming from design creators themselves and from others may be different (H5 is confirmed).

H5: Both more experienced and less experienced designers will be inaccurate in assessing their own performance level.

3.5 Conclusion

In this exploratory study, we attempted to investigate possible variances in design processes with respect to alternative generation (e.g. style of alternative generation, number of alternative genera-
tion), design outcomes (e.g. novelty and usability), and participant background (e.g. experiences). Both self-reported measures and ratings by an expert panel were used. The quantitative and qualitative analysis from this study may be limited in the following ways:

- A relatively low number of participants (n=18) may weaken the results for quantitative analysis. This sample number may be adequate for the exploratory nature of this work, and meets the standards for qualitative design studies, but needs further studies to make the conclusions stronger.

- Some may argue that even conceptual design for a website is a task that is too complicated for just 30 min. Subjects may employ strategies that are not natural to their normal design routines, and instead conform with the time-limit to deliver the best solution. For example, one participant did report that time constraints prevented him from even branching to other alternatives in his mind because he only had the time to work on one in detail.

- The use of Dreyfus’ model as a rubric could be questioned. During the process of navigating through the layers of the model to acquire expertise, the learner must abandon previously used methodology, patterns and cognitive strategies, and adopt new ones. Often, students get confused during the transition stages, and even experience suboptimal performance, as mental processes are completely overhauled. Hence, if participants were experiencing transitions, the results could be skewed.

Future studies should include more participants, more experts in the judging panel, and more robust evaluation criteria. We would like to continue to investigate different styles of alternative exploration and its relationships with different design outcomes and level of expertise (based on the Dreyfus’ Model).

Results from this study provide valuable insights to the different design processes and inspire interface design recommendations for computational design tools involving the same processes.
Chapter 4

Study 2: Site Interviews with Architects on Design Explorations and CDTs

4.1 Introduction

Computational design tools (CDTs) are getting ever more sophisticated and gaining traction in architectural design. The evolving technology impacts this practice with a rich experience and history from design processes and team dynamics in design firms. This study presents the results from six site interviews with design professionals in distinct design roles in design firms of varying nature. The goal of the interviews was to explore how computational tools are currently used in practice, especially in relation to alternatives generation and evaluation; and to identify existing challenges and directions for the improvement of future generative computational tools. We performed content analysis on data collected from semi-structured interviews. A combination of inductive and deductive inquiry is used in the investigation.

The new design tools mainly use parametric design models. To our knowledge, there is little research studying the effect of these tools on the task environment, and do not focus on the challenges that practicing designers are faced with. We hope that these interviews with practicing architects can shed some light on understanding, at least a small section of how computational tools are currently used in various design contexts. We believe these can help in analyzing the current challenges, assist with proposing guidelines for improving existing tools, and suggest possibilities for future research for generative computational design tools.

4.2 Methods

We conducted interviews as an exploratory and qualitative study to seek answers to the open-ended research question on how designers are coping with and absorbing computational design tools, into their design practice. We strived to select participants with diverse experience with computational tools; and we interviewed them on site, with the exception of the pilot run. Interviews on site
offer easier access to design artifacts when participants discuss specific projects, thereby providing a better understanding of the overall design context. Semi-structured interviews with follow-up questions, that may or may not be related to the pre-written scripts were asked at the discretion of the researcher. The interviews were recorded on images, video, and audio with the permission of the participants.

We transcribed the interview videos for analysis. An inductive open coding analysis was performed first to identify themes and subthemes with minimal presumptions on any themes. Then a deductive directed coding analysis was done to return to the structure of the original proposed list of questions. The combination of the two analyses, meet in the middle of the top-down and bottom-up approaches, and provide a summative qualitative analysis of our research questions regarding CDT use in exploring alternatives. Although efforts are made to minimize researchers’ own biases to the best of our ability, the results from this qualitative research are inevitably affected by factors like researcher views and interviewer field experience. We attempt to present our analysis process as transparent as possible to disclose such possible biases to the readers’ judgment.

4.2.1 Participants and Procedure

The interviews were conducted in the span of one month, starting with a pilot study with video conferencing tools. This was followed by five site visits (See table 4.1 for a summary of participant background). The data reported in this study removes individual names and necessary project details to protect participant’s privacy. The participants were recruited from personal or professional contacts of our lab members (Computational Design Lab @ SIAT SFU). There was no screening in participant recruitment, as our study was intended to survey the current practices in architectural design, as well their sentiments towards computational design tools and challenges they are facing in using the tools in work with alternatives or in their daily practice in general. In fact, there was a hope to have an interviewee pool as diverse as possible in terms of their background and experience.

4.2.2 Interview Structure

After the initial pilot study (P1), we refined the interview structure by reviewing details about time spent on projects and concrete step-by-step processes, using the questions in Table 4.2. This helped us to manage inconsistencies in the time and scope of the interviews. The pilot also demonstrated the difficulties of online interviews in sharing artifacts, so we decided to conduct in-person interviews instead. Another valuable outcome from the pilot was that we decided to leave the structure of the interview rather loose, and let the participants freely express themselves. This also enabled the participants to answer subsequent questions without interruption. The interviews ran from 45min to 2.5 hours.
Table 4.1: Summary of the interviewed participants (P) and firm background.

<table>
<thead>
<tr>
<th>Title</th>
<th>Education</th>
<th>Location</th>
<th>Firm Size</th>
<th>Experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 Principal</td>
<td>B. Arch &amp; M. Arch.</td>
<td>Lahore, Pakistan</td>
<td>&lt;= 5</td>
<td>5-10 yrs</td>
</tr>
<tr>
<td>P2 Architect</td>
<td>B. Arch &amp; M. Arch.</td>
<td>New York City, US</td>
<td>10+</td>
<td>5-10 yrs</td>
</tr>
<tr>
<td>P3 Developer (R&amp;D)</td>
<td>B. Arch &amp; M. Arch. M. Env. Bldg. Dsgn.</td>
<td>New York City, US</td>
<td>800+</td>
<td>10-20 yrs</td>
</tr>
<tr>
<td>P4 Senior Consultant</td>
<td>B. Arch Ph.D. Arch.</td>
<td>New York City, US</td>
<td>10+</td>
<td>5-10 yrs</td>
</tr>
<tr>
<td>P5 Principal</td>
<td>B. Arch &amp; M. Arch.</td>
<td>New York City, US</td>
<td>10+</td>
<td>40+ yrs</td>
</tr>
<tr>
<td>P6 Associate</td>
<td>B. Arch &amp; M. Arch.</td>
<td>Vancouver, Canada</td>
<td>1500+</td>
<td>10-20 yrs</td>
</tr>
</tbody>
</table>

4.2.3 Data Collection

Several forms of quantitative and qualitative data were collected during the interview, including audio and video recordings, and pictures of the subject’s artifacts. We focused on audio data and notes taken by the researcher during the interviews. A semi-auto transcription tool is built to assist with the transcription process (details on how the tool is built and its performance are discussed in Appendix B).

Inductive Open-Coding

After the transcripts were completed, the open-coding process started by reading them line-by-line, and performing one pass of the initial coding. All codes are then listed and organized into a hierarchical structure. To avoid the need to differentiate themes, subthemes, and codes, we decided to call all levels in the hierarchical structure codes. The organization of the codes include creating new higher level nodes to group some codes, consolidating similar codes, and adding new codes based on established hierarchy structure. With the new hierarchical coding structure, the transcript data was then re-coded. Ideally, this cycle of coding, revising scheme, then re-coding should continue a number of times until no more new codes emerge. However, in this study, it only took one full cycle to complete the coding. Most of these codes turned out to be converging with the deductive directed closed coding done separately, with only a few unexpected themes.

Deductive Directed Coding

The deductive directed coding process is conducted by listing all prewritten questions of the semi-structured interviews, then searching for answers in the video transcripts. Since the interviews were
rather unstructured, for reasons discussed in the procedure section, this directed coding process was necessary to ensure all questions were answered and summarized. A summary of the converged codes and themes from both inductive and deductive coding process is presented in the following section.

Table 4.2: Interview script. Clarification and follow-up questions are not included.

---

**Background Questions**

1. What is your title in the company?
2. What is your work experience? (years worked, industry, title, types of work performed)
3. What is your education level? Relevant training in design (years, level, school, area of studies)
4. How would you rate yourself on the expert / novice scale?

**Design Practice Questions**

1. Could you please describe a normal design flow at your company from initial client contact, to conception, implementation and delivery?
2. What is your role in these projects?
3. Could you reflect on two specific projects that you have done that are representative
4. Please describe the process of you or your colleagues for generate design alternatives? (Probe numbers, collaboration, parallel / sequential)
5. Please describe your experience with digital (computer/mobile) programs that may generate design alternatives.
   (a) Whether used such programs or not, if not would it be useful
   (b) How have you used it
   (c) Reasons for not using it, if not used regularly

**Design Process Questions**

1. When you do have a large number of alternatives presented to you (e.g. shows some pictures), what would be your likely approach?
2. Do you have visual preferences of layouts in any aspects of your design process?
3. What do you think is your most three challenges in design and design-support tools in general?
4. What do you think is lacking in current design support tools?
4.2.4 Ethical Considerations

All images and transcripts of audio recordings presented in this thesis are reproduced with participant consent. Special considerations have been made to protect the anonymity of the subjects or the firm that they are working for.

4.3 Observations and Findings

4.3.1 Style of Generation-Evaluation Cycles

In the first study (Chapter 5), we focused on describing alternative generation by whether it is breadth-first, depth-first, or a combination of both. However, in Study 2 (current chapter), we focus on the temporal cycles involving both generation and evaluation of the alternative solutions. This is commonly referred to as the “design cycle” in the literature, with variations in stage classifications, with the basic generation-evaluation nature staying the same across variations.

Results from this study reveals different styles of generation-evaluation cycles. More quantitative studies are needed to uncover the different effects each of these styles brings to design outcomes and process. A summarized version of different generation types:

- **generate-evaluate one by one**: designer generates and evaluates one alternative at a time. This mostly is related to a depth-first generation pattern, and whether more alternatives should be generated is determined by the evaluation of results from the last solution. However, there are case reports that partial evaluation may be done right after the generation of some of the aspects, or on all aspects, but no final decision is made after the evaluation. In this case, it can be of the depth, breadth, or a combination of the styles of generation.

- **generate-select-evaluate**: designer generates multiple alternatives, and selects one or more to be evaluated, then continues the generation-evaluation cycle. This style can often be associated with the *breadth-first-then-depth* generation method. The subject generates a few alternatives for the current level of *breadth*, selects one or more to evaluate, and continues in *depth* for the selected ones based on evaluation results (discussed in Related Works Section 2.2). However, this generation-evaluation pattern is reported to be present in variations of alternative generation styles. For example, for *depth-first-then-breadth* generation, participants continue to modify one alternative until the solution is no longer workable. They generate more alternatives sequentially. When no more alternatives are generated, evaluation starts.

- **generate-all-then-filter**: designer generates immediately possible alternatives, then uses filtering mechanisms (e.g. evaluation heuristics and grouping) to perform the final evaluation. This is commonly reported when a computer program generates a large number of alternatives for the designer.
Interview subjects report that generation and evaluation patterns vary with the nature of the project, and during the course of one project, different styles of alternative generation and evaluation may be mixed for different stages of the project.

4.3.2 Displaying Alternatives

One of the questions that we are interested in is how people currently display and interact with alternatives in their working environment. We believe learning about possible layouts of alternatives and interactions with them can help us to understand how designers can work with alternatives better, and consequently help us to build better CDTs to support these organic processes. We identified classifications for types of displaying alternatives as the following:

- **Layout alternatives side by side.** This is commonly seen in both digital and physical artifacts. Both vertical and horizontal spaces are used for the display. Physical vertical space examples are walls, cubicle dividers, and presentation boards, which are the most used vertical spaces for displays of 2D images, specs, and physical frit patterns. One surprising finding was that the vertical space can be used to display 3D building model alternatives with the ground plane anchored to the wall, as if they were 2D floor plans. Another unexpected pattern was that no grid display for alternatives was seen during all interviews. All alternatives were presented in a 1D line, either horizontally or vertically aligned. Speculations could be that the number of alternatives were normally a small number that could fit onto the physical 1D space and there was no need for a grid structure. Physical horizontal spaces are used exactly in the same way as physical vertical spaces, though hold more temporary pieces, due to easier access to change positions. In digital artifact forms, it was found interviewees often only do two at a time comparisons of two artifacts. We speculate that screen sizes are a lot more restricted than vertical wall spaces. Also, most software interaction affordances do allow easy rearrangements of alternative layouts with the mouse and screen setup found in most architectural offices.

- **Playing with alternatives in an animation.** Two participants (P3, P6) have mentioned making an animation to play through all alternatives possible as a way of presenting to the client. This was done mostly to demonstrate the potential of the parametric model as a vehicle of design. However, the animation video is not interactive for clients to meaningfully explore using parameter controls.

- **Stack alternatives and review one by one.** *Stack* here means putting the alternatives in stack-like structures. This could be a physical form of a structure, or an abstract data structure that exists digitally. Participants report it is a common process to review alternatives one by one linearly. The difference between this way of displaying alternatives with the animation method is that designers tend to spend more time on each alternative and manual control the pace of going through the “stack”.

62
4.3.3 How Many Alternatives? And Which Ones?

There is a big interest in learning how many alternatives designers work with, and how design decisions are made during the exploration process for which alternatives are selected. “How many alternatives” here is a vague question - we are interested in the order of magnitude, rather than a detailed range of numbers. From the interview study, we identified the following typical cases in terms of number of alternatives:

- **1 Alternative:** Often when resources are limited, in smaller firms or for small projects, the designers tend to fixate on 1 alternative. They produce additional alternatives only at a client’s request or when unresolvable problems surface during implementation.

  “Normally, we try to stick with one idea, and convince the client on that idea. But if it doesn’t work out. In that case, that’ll be a back and forth process between the client and us, and we probably devise the project 3 or 4 times to different options. sometimes recommendations are made by the clients. But this is how it works” – Quote P0

- **Less than 5 Alternatives:** Most of the projects mentioned during the interviews fall into this category. In general, 2 to 4 alternatives are generated and presented. This can either be for the entire project, or a specific aspect of the project in the design development.

- **Around 100 Alternatives:** This is typical in cases involving generative design tools or larger, more open-ended projects with more designers involved. Commonly a parametric design model is built with less than ten parameters, and the designer then selectively generates alternatives with only two or three selected parameters, each with only a few possible values. The resulted set of around 100 alternatives is then hand-picked by the designer. During this process, custom visualizations representing alternatives’ statistics may be used to aid with the selection process. However, most of the time, designers just manually go through the alternatives one by one on a computer screen, or even compile generated alternatives in an animation for both internal team meetings and external client presentations. As one exception of manual alternatives creation at this scale, P4 had one project that involved 10 designers over a period of 6 months and started with few constraints. In this project, each group members individually brainstormed ideas for alternatives, and presented their ideas at workshop critique sessions daily, learned from the alternatives, but threw them away, then repeated the process the next day.

- **200+ Alternatives:** At this scale, the process is typically reported to be for testing purposes only or for very early stage explorations. Usually, the alternatives are generated by a computational design specialist. In the interviews, this number ranged from hundreds to thousands. The exact number of alternative at this scale does not concern the user, rather it is merely treated as “a large number”, as no manual operations are directly performed. Different forms
of computational selection will happen to bring down the large scale to one of the three levels above (1, less than 5, or around 100 alternatives) and continue from there.

Note that, often detailed design development calls for design alternatives for sub-problems in the project (e.g. facade or staircase placement). We classify this kind of alternative generation as a modular design problem, where each problem-space is considered independently, i.e. not considered as a combinatorial design alternative possibility for a holistic design solution. For example, a designer is exploring different staircase placements and facade alternatives at the same time independently. There are normally no dependencies between the two problem-spaces, such that one staircase placement can only be applied for particular facade(s). Then, the number of alternatives generated from these sub-problems are perceived by designers individually, i.e. 3 facades and 4 staircase placements, as opposed to counting alternatives as if the sub-problems are integrated into a whole design, i.e. 12 building alternatives (3 facades x 4 staircases). Based on this perception from the designers themselves, the number scales we presented above account for a qualitative observation of all instances of different project stages, sub-problems, and iterations. For example, a design team may produce only 1 Schematic Design (SD) in one project, but during Design Development (DD) produces 3 room configurations, with 4 window designs. We consider these numbers individually.

We have also probed the exact design decision process involving these alternatives. Who makes the decisions and how? This is essential in understanding how to build support tools for designers to accommodate various numbers of alternatives and design decision making processes.

### 4.3.4 Use of Computational Design Tools (CDTs)

Participants have a wide range of experience on CDTs. In this study, we are particularly interested in participant behaviours during the SD and DD phases. However, how participants used a CDT in distinctly ways different from one and another. All participants sketched with pens or pencils during the initial phases. Some participants used other tools (e.g. physical massing forms, or CDT) to “sketch out ideas” and to present designs to colleagues and clients.

**Frequency of Use**

Ranging from use on a daily basis to once every project.

**Role in Creativity Support**

To understand CDT roles in creativity support in different scenarios, we took an analogy approach, to name each role with an object that mimics the same functionality. A designer can be using CDTs for more than one of its roles in the same project, or just a single role for an project.
Typewriter Role: Transfers hand-drawn sketches to formal drawings. One of the most important and first recognized benefits of CAD tools is the standardized format with clear legibility of drawings, in comparison to hand-drawn design sketches. A practice that is still prevalent in some firms is a “drafter” position, that is solely responsible for transferring hand-drawn plans to CAD drawings. There is no creativity involved in this role and the CAD tool acts much like a typewriter in earlier days, whose job is solely to transfer hand-written notes into more formal prints.

**Chronological position in design process** This role can take place at any point of the design process. For designers who are more used to designing with CAD tools, the typewriter role of the CAD tool tends to happen very early in the SD stage after crude concept sketches. After transferring the sketches into CAD models (e.g. SketchUp or AutoCAD), they continue the “sketching” process within the CAD tool. However, for other designers who are more used to sketching on paper, the typewriter role of CAD appears much later in the design process, after most of the design is finalized. The final hand sketch is then transferred to CAD drawings, where little design changes are made after.

**Person(s) executing CAD typewriter role.** Throughout the interviews, we identified three different process types in regard to the person(s) executing the CAD typewriter roles. We named the processes: the **lone designer** process, the **production line** process, and the **pair drawing** process (Table 4.3).

Table 4.3: Different processes classified by the person(s) involved in executing the CAD typewriter role.

<table>
<thead>
<tr>
<th>Process Type</th>
<th>Sketch Creator</th>
<th>CAD Creator</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lone Designer</td>
<td>Designer</td>
<td>Designer</td>
<td>The sketch creator transfers his/her own drawing into a CAD drawing.</td>
</tr>
<tr>
<td>Production Line</td>
<td>Designer</td>
<td>Drafter or another designer (often more junior)</td>
<td>The sketch creator is only responsible for designing the sketch, and asks someone else to transfer it to CAD drawing.</td>
</tr>
<tr>
<td>Pair Drawing</td>
<td>Designer</td>
<td>Designer with drafter or another designer</td>
<td>The sketch creator produces a sketch first, sits with another person who will be operating the CAD software, and produces the CAD together. The designer clarifies the sketch and ensures the CAD drawing executed is accurate to the intended plan.</td>
</tr>
</tbody>
</table>
Although in all three process types, the CAD tool is being used in the typewriter role to help
the sketch creator to translate hand-drawn sketches into CAD drawings, there is a possibility that
the executor is also using the CAD tool in other roles (to be discussed in the later sections). For
example, in the lone designer process, the designer may modify the sketch as he transfers it into
CAD, thereby not only copying content, but also creating content. The dynamic can be particularly
interesting in the production line and pair drawing processes, as it may involve collaborative design
as the sketch creator or the CAD executor interacts with the design plan during the “typewriting”
process.

**CAD artifacts produced within the typewriter role.** CAD artifacts produced within the type-
writer role can either be used for developing the design further, or to show the final results to the
stakeholders. The interim artifacts designers used to study the problem and explore design alterna-
tives are also shown to the client with the final solutions in order to demonstrate the process. For
the purpose of the following discussion, I differentiate interim artifacts and presentation artifacts
by whether the designers decide it to be a satisficing solution or merely as a discarded exploration
attempt. The naming has no relation to whether the artifacts are shown to clients or not in reality.

Three processes are identified in relation to how CAD artifacts are produced and used within
the typewriter role as interim artifacts:

- **Sketch to CAD drawing**
  Converting a hand-drawn sketch into a formal CAD drawing is the most typical use of a
  “typewriter role” for CAD tools.(Figure 4.1)

- **Sketch added to CAD drawing, then transferred to new CAD drawing**
  CAD drawings are printed during the design process for review, discussion, and iteration.
  Designers sometimes mark on the CAD prints with pen/pencil to specify design changes and
draw in more details or annotations (Figure 4.2). These marked CAD prints are a form of an
interim artifact that gets transferred to a new CAD drawing.

- **CAD drawing to new sketch**
  During the design process, designers sometimes use tracing paper to trace part of a printed
  CAD drawing as a way of making design revisions or creating new design alternatives (Figure
  4.3).

A designer may or may not use all three processes above during his design. Since design is
non-linear, these processes may repeat in cycles at various points of the design process.

The history of using semi-transparent tracing paper in design dates back as early as the 1800s.
The traditional design practice relies heavily on tracing paper. It is interesting to observe how this
practice is transitioning to interact with the use of CAD tools, rather than being replaced by similar
functionalities of CAD tools in this case. Some designers report to using CAD tools for revisions
Figure 4.1: Sketch to CAD. Adapted with images licensed under Creative Commons CC BY 2.0 © 2010 Kelvin Dickinson (Dickinson, 2010a,b).
Figure 4.2: Sketch + CAD = new CAD. Adapted with images ©2015 Bob Borson, reproduced with kind permission (Borson, 2015).

Figure 4.3: CAD + tracing paper = new sketch. Adapted with image ©2012 Bob Borson, reproduced with kind permission (Borson, 2012).
Designers use interim artifacts described above to refine their designs, until they produce presentation artifacts to illustrate solutions to clients. These artifacts may or may not be final, but are often used for communication with clients at milestone stages. I identified five types of presentation artifacts from interviewee reports:

- **2D non-digital renderings**
  Examples of these would be coloured pencil drawings or water-colour paintings. Only one interviewee out of the six used 2D hand-drawn renderings as a presentation artifact. The rest used 2D non-digital sketches only as interim artifacts to record design ideas and for internal communications (Figure 4.4).

![Figure 4.4: Examples of 2D: non-digital renderings.](a) A water-color rendering, licensed under Creative Commons CC BY 2.0 © 2011 Jose Javato (Javato, 2011)

![Figure 4.4: Examples of 2D: non-digital renderings.](b) A non-digital concept sketch, licensed under Creative Commons CC BY 2.0 © 2011 Jay@MorphoLA (Jay@MorphoLA, 2011c)

- **2D skeleton CAD drawings + non-digital media** Designers may use a skeleton CAD drawing that typically has incomplete details of the design, and use non-digital media to complete the representation as presentation artifacts (Figure 4.5).

- **2D detailed CAD drawings**
  The most common and standard form of presentation artifacts are detailed CAD drawings. A
detailed digital CAD drawing is preferred by many designers during the SD and DD stages because it has less ambiguities in the design (Figure 4.6). They can also reduce communication efforts during the CD stage, and the industry standard for construction is almost completely digitized.

diamond 3D digital renderings
Computer software can be used to render 3D models of the design, commonly with the addition of human, cars, furnitures and other environmental objects to render the design as realistically as possible (Figure 4.7). Sometimes, renderings are intended to be “photo-realistic” and can hardly be distinguished from real photographs of buildings. Sometimes, animations are created from these 3D models to give a “finished” touch.

diamond 3D physical models
Designers often build 3D physical models to study forms, explore options, and exhibit the final design to clients (Figure 4.8). It is useful especially in explaining complex or an unusual design. Designers report that for most clients, physical models are the most effective in visualizing the design from various angles when the design team presents.

**Sketcher Role: Functions like a pencil to capture design thoughts.** The *sketching role* of CAD tools allows the designer to use them like a pen to sketch their design. In this role, the designer is using the tool during the process of creating content, as opposed to copying existing content. When a designer is transferring a hand-sketch into CAD drawings (*typewriter role*), the *sketching role* often either resumes or occurs simultaneously.

Tool proficiency is found to be correlated with how much a designer uses CAD tools for sketching. One participant in particularly was fond of using non-digital media for design (e.g. pencil, pen, water colors). In most of his projects he uses only the typewriter functionality of CDTs and only uses CAD tools at the latest possible stages. He commented that using traditional sketching is “more flexible” for him, but CAD tools are “better at drawing straight lines”. His process is unique among all interviewees, but might be representative of designers who prefer non-digital media for both the creative process and final presentation.

**Joint Creator Role: Interacts like a human design partner.** The CAD tools can act like a design partner to the designer. Parametric capabilities from softwares like Grasshopper and Generative Components are capable of generating new designs for a given set of parameters. The tool can suggest design alternatives and join the designer in creating new solutions together. Detailed investigations about how designers behave in this role are described in Section 4.3.4.

**Spreadsheet Role: Provides or assists with automatic and real-time calculations for verification.** As interim design artifacts are being created during the design process, the designers constantly need to evaluate the feasibility of solutions against requirements. Traditionally, these can be done manually, but some CAD tools are capable of automating this process to speed up the
(a) CAD print + Markers, licensed under Creative Commons CC BY 2.0 © 2011 Jay@MorphoLA (Jay@MorphoLA, 2011b). Markers were used on a printed CAD skeleton drawing to create a more realistic renderings with colour, texture and shadows.

(b) Process of how the above sketch is created, adapted from image licensed under Creative Commons CC BY 2.0 © 2011 Jay@MorphoLA (Jay@MorphoLA, 2011a).

Figure 4.5: Example of 2D drawings with printed CAD drawings + non-digital media, and its creation process.
(a) Floor plan of a detailed CAD drawing, licensed under Creative Commons CC BY 2.0 © 2010 Zane Selvans (Selvans, 2010).

(b) Detailed floor plan of a building, licensed under Creative Commons CC BY 2.0 © 2015 Grey Emel (Emel, 2015).

Figure 4.6: Examples of 2D renderings with detailed CAD drawings.
Figure 4.7: Examples of photo-realistic 3D architectural rendering.
Figure 4.8: Examples of physical 3D models built during the design process.
design-evaluation cycle. For example, architectural designs require calculations on how environmental factors like sun, wind, or geothermal are affecting lighting and energy consumption of the proposed design alternatives. It may also verify designs against building codes. Designers can also use automated cost estimation to ensure proposed designs are within the budget.

Although this role is not directly related to creating designs, participants report that when familiarized with the spreadsheet role, their design exploration process changed significantly. Some of these calculations are time consuming when done manually and are often left until the later stages of design, after alternatives are narrowed down to a small set. Hence, without the spreadsheet-like functionality, designers either consciously or unconsciously stay well within the boundaries of “safe” design in order to reduce the chances of being eliminated later after calculations are done. A real-time feedback system that verifies the designs against constraints, or calculates desired parameters can reduce efforts and resources needed for manual checks, consequently encouraging more exploration on the designers’ behalf during initial conceptual design. The spreadsheet role of CAD tools expands designer’s exploration space by allowing them to quickly test early concepts without risking extra costs if the design is discarded after expensive calculations.

Characteristics of Design Exploration Process with Generative CDTs

Four of the six interviewed participants were using CDTs for generating designs. How these tools were used not only varied by the individual, but also varied depending on the project.

We categorized how CDTs are used in the design exploration process by a few characteristics that are orthogonal to each other. We then used these characteristics to describe typical scenarios in the interviewee’s design practice of how they used CDTs for explorations.

Stage of design Participants reported the use of generative CDT in all of the essential design stages: Schematic Design (SD), Design Development (DD) and Construction Document (CD).

Aspects of design Participants reported aspects of architectural design where they used generative design. Common uses for form finding were façade patterns and geometry structures.

Computational tool boundary The tool boundary is another characteristic for describing G-CDT use in design exploration. This boundary has three dimensions: the human and machine axis, the digital and non-digital axis, and the generation and verification axis. Where the tool sits on the three axe defines the boundary of the generative CDT for design exploration.

Parametric model creator and explorer Comparable to the relationship between type-writer and sketcher, the parametric model creator and explorer are two roles that can be played by the same person, although often, roles are played by different team members. The amount of training someone needs to become proficient in parametric modeling is more than that of a regular CAD program. In addition, parametric modeling is relatively new to architectural design and has not
been integrated into the modern architectural educational curriculum as extensively as basic CAD training.

Even at firms that recognize the value of parametric design and regularly use parametric tools, not all designers can create and proficiently use parametric models. There is usually a parametric specialist or a team of parametric specialists at the firm to assist other designers with the design exploration process. Some firms hire special consultancies to work with them to build such models.

**Parametric exploration** Parametric iteration has two levels of connotations, 1) iteration of parameter values to create design alternatives, and 2) iteration of the parametric model by changing the model structure and constraints. It is possible for a design process to have many parametric value iterations that creates numerous design alternatives, but do not have experience a single parametric model iteration.

**Parametric value iterations** After a parametric model is built for design exploration, there are many ways of producing parametric value iterations to create different design alternatives. The interviewee’s report that they use one or a combination of the following methods in creating parametric value iterations:

- **Manual parameter manipulations (single)**
  Designer manually manipulates parameters to create one alternative at a time. The most common form of their manipulation is through an interactive interface built by the parametric modeler as part of the design model. The designer uses sliders, text-boxes, and other controllers to change parameter values that will change a visual representation of the design in an interactive 3D view. The designer can change camera orientation, lighting, and other environmental parameters to inspect the generated alternative in detail. Another form of manual manipulation is through entering values into a spreadsheet or the program code itself, and then generate the design.
  An interesting case is observed when a parametric change produces a new solution, which in turn becomes input for another tool for a further analysis, such as for an energy analysis. The CDT here takes a high-coupled role for both the spreadsheet role and the joint-creator role. For example, P4 sets up energy analysis calculations for any geometry created in the CDT site environment to perform real-time sunlight and geothermal calculations for all days in a year. He then creates different masses on the site as design alternatives while getting real-time analysis feedback. We classify this type of interaction as a manual parameter manipulation for creating one alternative at a time.

- **Automatic parameter generation (single)**
  During parametric value iterations, some parametric models have the functionalities of generating one alternative at a time without users directly specifying individual parameters. Interviewee’s report that when building parametric models, sometimes an “I’m feeling lucky”
function is requested by the client for their explorations. Other technology, like machine learning and genetic algorithms, are used in creating the next design based on existing preferences.

- **Automatic parameter generation (batch)**
  In particular projects, designers do ask the CDT to generate an uninterrupted batch of alternatives. Only one interviewee reported the use of parameter value iterations in this category. The batch values were specified by parameter intervals and a pre-determined step size for each parameter. The batch of alternatives is exhaustive within the range given the step size.

- **Parametric model iterations**
  In the absence of any digital or parametric tools, designers constantly reframe the design problem and constraints as they progress in the design process. If this were to transfer to the use of generative CDTs, one would expect that the parametric models are constantly refined to reflect the same process. However, I received mixed feedback on this aspect.

One participant worked in a consulting role for a project to create a parametric model of a high-rise residential building for a client. The model is built with Grasshopper and Rhino, with various sliders and text boxes for user input to control the geometrical attributes of the building. The participant commented that "I built it based on client's requirements, gave it to them, and never heard back from them again." There is also the sentiment of using the design model as a throw-away artifact to learn from, and build new models each time rather than reiterating an old one in conceptual stages: "I would say the parametric model is not persistent through the design stages. Maybe it's like a sketch model, something you change quickly. But yea... It's not like something you start on day one and on day hundred you're still working on the same model".

**Scenarios of Design Exploration Process with Generative CDTs**
In this section, I will describe several CDT scenarios surfaced during the interviews, with the above classifications for CDT use in design exploration process.

### 4.4 Conclusion

This study presents results from interviews with designers on how computational design tools (CDTs) are used in architectural practice by identifying codes and themes emerged in the analysis. Based on this, we propose a classification of characteristics of generative CDT use and discuss key challenges and guidelines for future CDT design.

This work inherits the limitations of a typical qualitative interview study, especially with only one data collector conducting the interviews and performing the analysis. Inexperienced with conducting semi-structured interview and analysis, best efforts were given towards familiarizing myself with the domain, not asking leading questions, and being transparent with my research intentions in the analysis and writing. An open-mind was kept during the investigation, and many unexpected
findings emerged in the process. Another limitation is biases in sampling and generalization. One unexpected suggestion from one of the participants was to move the architectural industry towards more open source design strategies following the software industry, rather than to keep following the traditional architectural practice of keeping “secret knowledge”. We hope sharing the results of this study and follow-up studies can contribute to the growing open-source generative design research and CDT development community. Another possible bias is in the participant recruitment process. Although I intended to include as much diversity in the interview pool as possible, there are still biases in sampling from the people that I have easy access to. The results from this study may not be entirely representative of the general architectural design community, but may be a good snapshot of it. The conclusions are carefully drawn with full recognition of these limitations.

Findings from this study identified some of the different roles and challenges of CDTs in the design process. Design recommendations for developing CDTs based on this study are discussed in detail in 7. The next chapter (Chapter 5) continues to investigate design exploration behaviours via an observational study in a lab setting on an architectural design task with CDT generated alternatives.
Table 4.4: Process characteristics of Scenario A: Museum Project

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>In This Scenario...</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Role</td>
<td>Architect Designer</td>
<td>Works with a team of architects and the principal to complete the design across multiple offices.</td>
</tr>
</tbody>
</table>
| CDT Roles                    | Typewriter Role     | **Chronological:** Early in the SD phase.  
**Person(s):** The sketch creator (designer) is the CAD program executor that transfers the hand-sketch into CAD drawing.  
**CAD Artifacts:** presentation artifacts are detailed CAD site plans, floor plans, 3D CAD renderings, and 3D physical model. |
| Sketcher Role                | After the initial typewriting task, the interviewee continued to use a CAD tool as a sketching tool to modify the design as the main designer responsible for the project. |
| Joint Creator Role           | At the DD stage, participant worked with a Grasshopper specialist in their office to create parametric models for a facade pattern. |
| Generative CDT Use           | Stage of Design     | Only in Design Development (DD) stage.  
Used to seek for a complex facade pattern. |
|                             | Aspects of Design   |                                                                          |
|                             | Tool Boundaries     | **Human-machine:** Alternatives generated by machine.  
**Digital-physical:** Digital artifacts are produced from the model, but actual decision-making are done on printed physical representations.  
**Generation-verification:** Used only for generation. |
|                             | Parametric Model    | Model created by a parametric specialist with inputs of the team. No one else on the team had direct interactions with the model. |
|                             | Creation            |                                                                          |
|                             | Parametric Explorations | **Value iterations:** Manual parameter manipulations (single). With multiple iteration cycles, each with 3-5 alternatives branching off from one preferred alternative in the previous iteration cycle.  
**Model iterations:** None. |
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>In This Scenario...</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Role</td>
<td>Parametric Consultant</td>
<td>Works with a client to provide an interactive parametric model for client’s designers to explore design alternatives of a high-rise residential building.</td>
</tr>
<tr>
<td>CDT Roles</td>
<td>Sketcher Role</td>
<td>The subject is using the CAD tool to sketch out an interactive parametric model.</td>
</tr>
<tr>
<td></td>
<td>Joint Creator Role</td>
<td>The model is intended to jointly created with the client in the design process. The parametric model is meant to be versatile in assisting the clients to explore possible alternatives.</td>
</tr>
<tr>
<td>Generative CDT Use</td>
<td>Stage of Design</td>
<td>Pre-Schematic Design (SD) Phase.</td>
</tr>
<tr>
<td>Aspects of Design</td>
<td></td>
<td>Form-finding for building structure.</td>
</tr>
<tr>
<td>Tool Boundaries</td>
<td>Human-machine: Alternatives generated by machine.</td>
<td>Digital-physical: Subject unclear about actual model use.</td>
</tr>
<tr>
<td></td>
<td>Generation-verification: Only for generation.</td>
<td></td>
</tr>
<tr>
<td>Parametric Model</td>
<td>Creation</td>
<td>Model created by a parametric consultant and used by the client. No feedback or revisions were done on the model after its creation.</td>
</tr>
<tr>
<td>Parametric Explorations</td>
<td>Value iterations: Manual parameter manipulations (single). The model provides manual manipulation of one set of values at a time.</td>
<td>Model iterations: None.</td>
</tr>
</tbody>
</table>
Table 4.6: Process characteristics of Scenario C: Cultural Centre Project

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>In This Scenario...</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Role</td>
<td>Parametric Consultant</td>
<td>Works with an internal designer to provide energy and sunlight analysis of an existing cultural centre</td>
</tr>
<tr>
<td>CDT Roles</td>
<td>Spreadsheet Role</td>
<td>Computes and visualizes energy and sunlight analysis data of the year cycle at various times of the day. Individual computations were represented by a heatmap equivalent projected on a 3D model of the building. An animation was created to showcase all computations, with each individual frame representing a data point.</td>
</tr>
<tr>
<td>Generative CDT Use</td>
<td>Stage of Design</td>
<td>After Design Development (DD) stage.</td>
</tr>
<tr>
<td></td>
<td>Aspects of Design</td>
<td>Use to verify a mostly fixed design for energy and sunlight analysis.</td>
</tr>
<tr>
<td></td>
<td>Tool Boundaries</td>
<td><strong>Human-machine:</strong> Calculations generated by machine. <strong>Digital-physical:</strong> Only digital artifacts are produced from the model. <strong>Generation-verification:</strong> Used only for verification.</td>
</tr>
<tr>
<td></td>
<td>Parametric Model Creation</td>
<td>Model created by a parametric specialist with inputs of the designer. No one other than the parametric specialist had direct interactions with the model.</td>
</tr>
<tr>
<td></td>
<td>Parametric Explorations</td>
<td><strong>Value iterations:</strong> Bulk value iterations with fixed parameter values. The process of generating value iterations carried out without human intervention after it started with defined values. <strong>Model iterations:</strong> None.</td>
</tr>
</tbody>
</table>
Table 4.7: Process characteristics of Scenario D: Mix-Use Development

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>In This Scenario...</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject Role</td>
<td>Lead Architect</td>
<td>Leads a team of five designers to design a mixed-use development that includes residential, retail, and office uses.</td>
</tr>
<tr>
<td>CDT Roles</td>
<td>Spreadsheet Role</td>
<td>Parametrically generate and render building designs with flexible constraint parameters including ratio of residential/commercial and investment return rates based on leasing the property. Exemplar alternatives are hand-picked by the designer team, and then built into a show case reel to present to the client.</td>
</tr>
<tr>
<td>Generative CDT Use</td>
<td>Stage of Design</td>
<td>Pre-Schematic Design (SD) stage.</td>
</tr>
<tr>
<td></td>
<td>Aspects of Design</td>
<td>To establish parametric models and render alternatives with drastically different function ratios and structure.</td>
</tr>
<tr>
<td>Tool Boundaries</td>
<td>Human-machine:</td>
<td>Alternatives generated by parametric models, but hand-picked by designer for decision.</td>
</tr>
<tr>
<td></td>
<td>Digital-physical:</td>
<td>Only digital artifacts are produced from the model.</td>
</tr>
<tr>
<td></td>
<td>Generation-verification:</td>
<td>Used only for generation.</td>
</tr>
<tr>
<td>Parametric Model</td>
<td>Creation</td>
<td>Model created by a parametric specialist with inputs from the designer team. No one other than the parametric specialist had direct interactions with the model. Lead designer makes final decision after group meetings.</td>
</tr>
<tr>
<td>Parametric Explorations</td>
<td>Value iterations:</td>
<td>Some manual value iterations were done based on generated alternatives during selection phase for final client presentation.</td>
</tr>
<tr>
<td></td>
<td>Model iterations:</td>
<td>None.</td>
</tr>
</tbody>
</table>
Chapter 5

Study 3: Design Exploration with CDT Generated Alternatives

5.1 Introduction

In this study we are interested in studying the decision-making strategies and patterns used by designers, when they are presented with a large number of alternatives. We define large here as in a parametric space generated by a computer CAD program, which can be over hundreds of thousands of alternatives, if not millions. Past research in decision-making literature (see discussions in Chapter 2) is generally focused on a small number of alternatives (in our frame of reference to large), which normally involves less than 10 alternatives, but ranges from 3 to 30 in the literature reviewed so far in decision-making from domains including finance, personal or emotional studies and other technical fields such as fire-fighting, and medicine. In the design domain, most of the studies relating to decision-making are qualitative case studies with few participants, many with less than 5 in the literature reviewed so far, and do not involve a large number of computer generated solutions. In order to fully utilize the potentials provided by computational design tools, we need to study the decision-making strategies and patterns used by designers in the context of large numbers of alternatives. We hope this may inform building design decision-support tools to facilitate the alternative exploration and selection process more effectively.

5.1.1 Research Questions

The intent of this mixed-method study is to learn about decision-making strategies and patterns used by designers when presented with large numbers of alternatives, during design exploration. Specifically, the following questions were investigated:

- How do designers explore alternatives when presented with a large set of computer-generated alternatives?
What are the differences between decision-making strategies used by designers, if there is more than one strategy observed?

Our aim is to make interface design recommendations for computational design tools to support design exploration and decision-making behaviours observed in the study.

5.1.2 Coding Scheme Development Attempt

One of the original goals of this study was to develop a draft version of a reusable coding scheme for future quantitative studies for studying alternative exploration behaviours. Presumably, this coding scheme could be used for future computational design tool evaluations in the same area (design alternatives exploration). Due to time constraints, the coding scheme is still under refinement to achieve more robust inter-rater reliability for generalization. In this study, we present the initial efforts and directions of the coding scheme in development, as well as preliminary findings of designer behaviours in an overloaded alternative exploration task. An interactive visualization was built to assist the researchers to analyze coding results and seek patterns in participant behaviour.

5.2 Methods

This was an observational lab study based on both quantitative and qualitative methods to explore the design decision making phenomena and categorizing behaviours. The study analysis was done on data collected from questionnaires, in-depth interviews, and video recordings of the participants’ tasks.

5.2.1 Participants

The observational study included 10 participants, of which one was a pilot subject. We expected to recruit all expert architect designers with experience in parametric CAD tools for design exploration. However, due to lack of immediate availability for such participants (as discussed in Chapter 1), we have resorted to recruiting for a partial experience match. The make up of the participant experience is shown in the Venn diagram (Figure 5.1).

5.2.2 Task

Participants were given a set of 1000 computer generated CAD alternatives printed on 4”x6” unruled index cards. The set of 1000 were generated from a parametric model in Generative Components using methods described in Appendix C. The set intended to provide a visually diverse set of alternatives to the participants, as shown in samples in Figure 5.2.

Participants were asked to complete an architectural design task as shown in Table 5.1. The task description was deliberately written to be open to interpretation in order to mimic the ill-defined nature of most design problems. This ensures that there are unknown factors to participants, and
they have to make their own decisions. In real life settings, designers may add information that may turn more unknowns into knowns. However, more or less parts of the problems always remain ill-defined throughout the design process. The number of alternatives is chosen as an arbitrarily large number to ensure the participants are over-loaded with choices. In real life settings, designers may be over-loaded with a different number of choices, influenced by different task environments, different use of tools, and of course individual differences as well. In this study, participants are limited to using alternatives printed on index cards, pen and paper, and whiteboards. The task is also limited by a soft time constraint. Participants were told to expect the average task time to be 1.5 to 2 hours, but they could spend as much or as little time as they wished to. The actual time spent on this task by 10 participants ranged between 0.5 to 3 hours.

5.2.3 Procedures

The study was conducted in a small meeting room with whiteboards with one participant each time (as shown in Figure 5.3). There were a total of four cameras recording user actions during the session. Two digital cameras (red stars) were aimed to capture user actions around the two white boards. GoPro1 (blue star) was aimed to capture an aerial view of the table, mounted on the side table raised to about 70cm over the table top. GoPro2 was mounted on the participant’s forehead via a headstrap in an attempt to get a general proximity of the user’s attention. One or two observers were present during each session, sitting in the corners of the room to be as unintrusive as possible. The table was angled to give ease of access to both white boards.

The study proceeded in the following order of events:
The researcher briefed the participant regarding the study, which included the need to vocalize what is on the participant’s mind while performing the task.

The participant was asked to do a pre-task questionnaire (Appendix D), which included background questions about their design experiences.

The participant was asked to read over the description of the design context and design task (Table 5.1).

The researcher answered any questions the participant may have, and continued to answer them during the course of the study.

The researcher handed all 1000 index cards, in a randomly ordered stack, to the participant.

The participant performed the task (Figure 5.4), and notified the researcher verbally when the participant decides the task has been completed.

The researcher confirmed the participant’s final decisions on the index cards or other sketches they created.
### Design Context

The famous architectural firm Cubis located in Vancouver has been commissioned to design a multi-storey residential apartment in a relatively posh area surrounded by parks and close to downtown Vancouver. Inspired from the Cubism style and colors from the De Stijl, the architects in the firm have generated a set of 1,000 conceptual design alternatives using a parametric design tool. The three primary architectural elements common across all designs are:

1. Gray cubes representing residential units that each unit can be subdivided in to different spaces, or cubes can be combined to create larger units.
2. Red vertical bars representing structural columns and possibly accommodating vertical elements, which may or may not include circulation, staircases and elevators, open atriums, cross-leveled apartment units etc.
3. Green horizontal elements representing open spaces between residential, which may or may not include balconies, open terraces, common areas, parking lots etc. These can either be private spaces for each residential unit or as public decks shared by different units.

### Design Task

Your task as the architectural design consultant is to go through the designs alternatives given to you and select the potential solutions using your own methods and strategies. The solutions you select for further development must consider the functional and stylistic requirements of the project. Again the functional requirements are loosely defined in the three architectural elements. You are welcome to propose additional guidelines and design requirements you would want this building design to consider.

- The researcher follow-up with a structured post-task questionnaire (Appendix D) and unstructured follow-up questions.

### 5.2.4 Data Collection and Analysis

Various forms of raw data were collected during the sessions:

- Pre-task questionnaires completed by participants just before the task starts.
- Physical artifacts produced by participants while performing the tasks (e.g. sketches, notes, markings on index cards, and white boards).
- Post-task questionnaire orally answered by participants with notes by researchers.
Figure 5.3: Observational Study Room Setup.

- Video recordings (4 camera angles) of participants task process and post-task interviews.

Qualitative analysis based on the recursive abstraction research method was done for video data, cross-references with researcher’s notes taken during the study, and throughout the analysis process. The researcher followed an open coding procedure, to look for patterns in design exploration activities. Detailed coding at the action level was performed for P1 only (discussed in the next section) and analyzed both qualitatively and quantitatively, with assistance from an interactive visualization developed for this study. Descriptive quantitative data was also reported based on observation counts and questionnaire data. However, due to the nature of the study, there was a relatively low subject count and no meaningful inferential statistics were intended to be performed. The data analysis reveals observable patterns in design exploration process, as well as leads to coding scheme development for follow-up studies.
5.2.5 Design Process Visualization

A detailed analysis was done on P1 as a trial to study how further quantitative analysis to this data may reveal valuable patterns. The entire video was transcribed and time-coded for card movement in the participant’s mental space, namely for unviewed, speculating, or rejected. Every time the participant comes in contact with one card with his hand, an event was recorded for the card interaction. Additional free form comments are made to events that are not captured by the codes. Important participant quotes are also recorded. After the initial round of coding, the researcher went back to the coded events and code at a meta-level for evidence of repeatable cycles, criteria development, and other cognitive activities. The coding scheme was refined as more patterns were discovered, and iteratively re-coded to achieve internal consistency. Figure 5.5 illustrates a selected sample from the spreadsheet used for coding. An interactive web-based visualization (Figure 5.6) was developed as an analysis tool, to assist the researchers to see patterns in coding data. D3.js and JQuery Javascript libraries were used along with HTML/CSS.

Figure 5.5: A partial view from the spreadsheet showing all coded events from P1’s videos.
Figure 5.6: An overview screenshot of the design process visualization in basic mode. The **x-axis** represents the entire study session from the start (0:0:0) to the end (1:46:46). Each line running horizontally is a design alternative (card) that the participant has interacted with during the session. The **y-axis** is roughly separated into three areas, a) unviewed cards, b) cards being viewed or accepted temporarily or permanently, c) rejected. The visualization at this zoom-level is used for viewing global process patterns. Zoomed-in views are shown in Figure 5.7.

The researcher may interact with the visualization through zooming, mouse hovering, and clicking (Figure 5.7). The visualization can show all coded events of an individual alternative over time. For example, Figure 5.8 shows details of two selected design alternatives. It is evident that the participant spent more time viewing and deciding about the top alternative in the earlier session, but only spent a fraction of the time about the lower alternative as decision criteria become more concrete and clear.

More advanced interactions are developed on top of this original basic mode to display events or patterns with more context. For example, he could involve vertically phasing lines based on specific event tags (e.g. start or end of a decision cycle), and showing one or more type of event durations across the session (e.g. all events related to design criteria creation, ranking, and removal). These will be discussed below in the results section in more details.

### 5.3 Results

Participants exhibited different patterns throughout the sessions. We attempted to capture their design decision-making behaviours under a predefined design task scenario. The scenario was based on:

- A large number (overloaded) of computer generated alternatives were presented to a designer for a specific design task for design exploration.
- The designer explored the alternatives was unaware of how the alternatives were generated.
- For this specific phase of exploration, no interaction with the computer alternative generation mechanism was possible.

From Study 2 (Chapter 4), we learned that similar situations like this study do happen often in user roles of **generative model designer**, **generative model value iterator**, and **viewer**, and are currently now well supported. Often interactions with the CDT is either limited by the user’s ability or knowledge of the CDT in the **viewer** role. And commonly, users in **generative model designer**
Figure 5.7: Zoom-in views of various interaction details of the process visualization.

1) Hovering over a line
   Highlights line in red

2) Clicking on a line
   Shows coded events as circles

3) Hovering over a circle (an event)
   Shows details about an coded event.

Grey Circles
Events were only coded for card state change. Only a card number representing the design alternative, and the timestamp of the event are displayed.

White Circles
Events were coded for more details including comments from coder, and participant quotes during the event.

Figure 5.8: Examples of two design alternatives visualized clearly showing different level of interactions from the participant during the study session.

A card that went through multiple events of inspection and state changes. Constantly being viewed until eventually discarded.

A card that was only briefly looked at by the subject. It was quickly viewed and discarded.
or value iterator are conducting a speculative exploration for this specific iteration (for model or value) before the next iteration. By generalizing user behaviours under these assumptions, we aim to suggest interface design recommendations for future generative design tools.

A generalized description for user decision-making behaviours during design exploration is a cycle (Figure 5.9), and users appeared to follow these steps in the exploration process:

Figure 5.9: Observed cycle for design exploration with a large number of computer generated alternatives. Numbers in the graph represent steps in the cycle, corresponding to detailed descriptions in the text.

1. Initial strategy developments, observed in three modes:
   - **One at a time, continuously**: participant scanned or viewed alternatives one by one, and typically made a decision after viewing each individual alternative.
   - **One set at a time, simultaneously**: participant speculated alternatives as a set simultaneously by laying cards out on a table.
   - **A mix of both**: one at a time and one set at a time.

2. Develop criteria for filtering alternatives
   - Types of filtering observed:
     - **Additive selection**: multiple criteria were evaluated as a whole depending on importance weighting for each criterion (e.g. the use of heuristics).
     - **Elimination**: if a criteria is not passed, the alternative is eliminated regardless of evaluation results from the other criteria.
– **A mix of both**: additive selection and elimination.

- Classification of criteria observed:
  
  - **Visual (aesthetics) vs. Structural vs. Functional**
    A continuous scale of whether a criteria represented more visual, structural, or functional aspects of the design.
    e.g. too boxy (visual aesthetics), not enough ventilation (functional).
  
  - **Physically recorded vs. Non-physical**
    A binary value of whether the criteria was physically recorded or not.
    e.g. written on a post-it note (physical), grouped on table or board (physical), spoken (nonphysical).
  
  - **Explicit vs. Implicit**
    A binary value of whether the criteria was explicitly specified through verbal or writing. This is a different dimension from physical recording.
    e.g. a verbalized statement like “not tall enough” indicated a criteria for building height, which is non-physical and explicit.
    e.g. grouping some cards together on the table based on similar forms looking like each other without any verbal or written specification on the criteria of selection, was both physical and implicit.
  
  - **Well-defined vs. Ill-defined**
    If the criteria rule was defined so that anybody who evaluates it renders the same result then we call it well-defined or objective.
    e.g. number of floors in the building measured in scale (well), balanced colours of red/green/grey (well), visually pleasing (ill), earthquake proof (ill in this context due to difference in knowledge of earthquake standards).

3. Filter alternatives with developed criteria, may go back to the last step for more criteria development. This step can include the following (can be simultaneous actions) other than individual additive selection or eliminations:

  - **Grouping**: grouping similar alternatives together based on a criteria for further selection or elimination actions.
  
  - **Ranking**: ranking alternatives by a criteria, then select the top few or eliminate the bottom few.
  
  - **Comparison**: comparing two or more alternatives side by side for one or more criteria for further selection or elimination actions.

4. Reflection on resulted smaller set of alternatives. May modify or create new alternatives based on resulting set.

5. (Optional) Go back to step 1 or 2, and repeat cycle as much as needed.
6. [Goal State for Exploration] Set design guidelines for next stage, with different possibilities in observed outcomes:

- Presented one or more alternatives selected, with no modifications (Picture 6, Figure 5.16).
- Presented both selected and eliminated alternatives as example designs, with no modifications.
- Presented modifications based on one alternative, to be further developed alone (Figure 5.10).

Figure 5.10: P9 and P5’s final deliverables for design exploration task, with both based on a single alternative.

- Presented modifications based on multiple alternatives individually to be further developed in parallel.
- Combined multiple alternatives into a new alternative, to be further developed alone (Figure 5.11).

Figure 5.11: P6 and P7’s final deliverables for design exploration task. Both based on inspirations from multiple alternatives seen, but combined with the designer’s own elements, result in a final design that did not resemble any of the alternatives in the generated set.

- Combined multiple alternatives into multiple new alternatives, to be further developed in parallel (Figure 5.12).
Figure 5.12: P4’s final deliverables for design exploration task. Two alternatives produced with different views (top, front, back, side) are based on inspirations from multiple alternatives seen, and combined with designer’s own elements, resulting in a final design that did not resemble any of the alternatives in the generated set.
The above generalized design exploration cycle was observed in all participants. Three participants representative of different styles of explorations are described below to show how the abstraction became distilled from these observations:

**P2** This participant went through 3 cycles of design exploration, procedurally performed an initial elimination on the table based on all criteria developed so far, then put selected cards onto the board for further grouping and elimination, until the final contenders are found.

0 min: He speculated alternatives for initial strategy one set at a time spontaneously (Figure 5.13), by laying all cards on the table. He first used explicit verbal (non-physically recorded) elimination strategies (e.g. not enough spaces, not structurally sound, not making sense) for mostly structural criteria. He then used implicit ill-defined criteria (unclear to the researcher) to select a smaller set (4 cards) from the table to be put up on the board.

Figure 5.13: P2 viewing alternatives one set at the time, attempting to develop criteria.

13 min: He started a new cycle, changed to speculating alternatives one at a time continuously (Figure 5.14), sorted cards into yes and no piles with criteria developed in last cycle. He then put up all yes cards onto the board (total 25 cards on board), and grouped alternatives by explicit visual criteria (e.g. this one looks funny, these are complex, 2 separate buildings vs. 1 building), while continuing to eliminate 6 alternatives based on explicit structural criteria.

Figure 5.14: P2 viewing alternatives one at a time using only non-physical recorded verbal criteria.

28 min: New cycle started. Speculation patterns were the same as last cycle. For the first 35 min of the session, the participant used mostly non-physically recorded verbal criteria...
(except physical grouping on board), both implicit and explicit. He developed explicit written criteria on another white board (Figure 5.15), mostly visual with one structural. Cycle continued with the same actions as the last cycle, sorting yes/no piles, and putting selected on board while eliminating some, mostly based on same grouping visual criteria. Picture 1, in Figure 5.16 shows the board state at the end of this cycle.

![Figure 5.15: P2 Design Exploration Process.](image)

58 min: New cycle started. Speculation patterns remained the same. When putting new cards onto the board (at 1 h 25 min), he started to create new groups based on visual criteria, and explicitly recorded the criteria on the white board (Picture 2, Figure 5.16). More criteria were created as he continued the grouping and elimination process. (Picture 3 to 4, Figure 5.16) based on visual composition, including 1 mass, 2 mass vertical, 2 mass horizontal, 3 mass vertical, 2x2 H&V, 2x2x2 H&V, complex, interesting. In Picture 5, Figure 5.16 (at 1 hr 50 min), 2x2x2 H&V was erased to make room for the final selection category contenders (Picture 6, Figure 5.16).

1 h 58 min: End of task.

**P4** This participant went through 2 cycles in the design exploration with generated alternatives, spending most of his time developing his own alternative sketches without viewing additional new cards.

Cycle 1: He spent the first 4.5 min of the session in a quick Cycle 1 speculating the 19 alternatives as a set, and developed criteria for selection and elimination mostly implicit ill-defined, such as “because there’s not enough to work with”, or “might be interesting”. The block quote below shows some more examples. After viewing cards with the criteria (Figure 5.17, left picture), left 3 selected for the next cycle.

“At this point, I’m just trying to use these structures you’ve got here to understand what the problem is and how the firm has already approached it. At this point, I’m still very uncertain, as a consultant, what I’m supposed to be doing. So I’m trying to figure out based on these, what the idea is [...] I got rid of the bits that didn’t
have residential units in them. I’m also getting rid of the structural parts that don’t seem to have any purpose. I don’t know. This is kinda of the instinct that I’m going with right now... That one I just don’t like the look of it 1[...] At this point, I’m just eliminating a bunch based on lack of visual interest. This one, structurally doesn’t seem to make any sense to me.. neither is that one. [...] This one looks out of balance...the balance isn’t right, looks like there’s a lot of residential area that it can’t support it. these two are very similar, but this one seems to be better than the other two. And these three I’m just getting rid of, because comparing to the other ones here at the moment we’ve got, there’s not enough to work with.” – P4 quotes during cycle 1 criteria development

Cycle 2: The first steps of speculation (Step 2) and filtering alternatives (Step 3) only lasted less than 2 min. P4 quickly scanned through 43 cards randomly picked from the unviewed pile. He selected another 3 cards based on the mostly implicit and ill-defined criteria developed in the last cycle. A total of 6 cards were selected as inspirations for the next steps.

“At this point, I have a better idea of what kinds of things the computer is generating, I’m starting to get a better idea of things that I find interesting. At this
Figure 5.17: P4 Design Exploration Process. Left: Speculating on cards one set at a time initially. Right: Sketching original solutions based on inspirations from the cards.

point, I’m going through much more quickly, because at the stage of the process, I’m looking for novelty, rather than any kind of a solution in the cards. [...] At this point, I think I have gone through about as much card as it will be useful, in terms of making ideas and guiding things. I’ve got a number of cards here are very different with each other, but each one of them have particular things in them that I like. At this point, it seems like starting to move away from the cards and onto paper is the best idea, the most constructive idea.”

– P4 quotes at the end of Step 3 for Cycle 2

After that, he spent the rest of his 48 min session sketching and developing his own alternatives, without viewing any more new cards (Figure 5.17, right picture). Some of the interim sketches are shown in Figure 5.18, and final sketches presented at the end of the task are shown in 5.12.

“...the style that’s been chosen and the [computer] drawings don’t match. There is a playfulness to the [cubism] style that isn’t present in the [computer drawings] here. And the next stage in whatever the process is for the company would be getting some of the stylistic elements and the playfulness back into it. With that in mind, I put together a couple of concepts, that might be able to be used... ”

–P4 quotes for presenting final alternative designs sketched

P6 This participant goes through 2 cycles of design exploration. With a mostly explicitly well-defined set of criteria, she commented that for the next stage of design exploration she would like to have a junior designer or a computer to continue the exploration by performing the procedures she established.

Cycle 1: In the first 16 mins of the session, she speculated on the alternatives one by one, and developed explicit and physically recorded (by grouping placement on table) criteria incrementally. Picture 1 to 4 from Figure 5.19 shows the development of 8 criteria in total. All filtering was done based on elimination strategies simultaneously. At the end of the cycle, all remaining alternatives are inspected again to catch ones that she missed. P6 is different
than most other participants in that she kept a classification of the rejected alternatives, and used it in her final presentation of results to show “what not to do” along with the selected alternatives. She also commented on the flexibility of thresholds in parameters that may change her decision outcomes. For example, she grouped buildings by height into 3 groups: low, mid and high rises. She assumed what she wanted now was mid-rises, but keeping the group of the rejected high-rises allowed her to include it if criteria threshold changed.

Cycle 2: Continued in the same pattern as in Cycle 1 using the same criteria. No more new criteria were created until the last step of reflection (Step 4). She regrouped the final selection based on visual criteria (e.g. two mass buildings, stilts buildings) while eliminating some that she missed in the first pass. During this process, she developed a new criteria (Picture 5, Figure 5.19 for a green balcony that is too big and can block light to all floors below. She finally ended up with 6 alternatives at the end of the task, all filtered through explicit elimination at 29 min mark of the session.)

The selected detail descriptions of design exploration cycles demonstrates at the generalized steps may be an accurate abstraction of the design exploration cycle with computer generated alternatives. The next few subsections of the results section will illustrate other findings of the design exploration process.

5.3.1 Self Descriptions of Design Exploration Process

In the last question of the post-task questionnaire, participants were asked to describe their design exploration and decision making process to their fellow design colleagues working on the project on a sheet of paper, using their choice of representation. Three types of representations were found among the participants: abstract visual, detailed visual, and procedural textual. How partici-
Figure 5.19: P6 design exploration process for Cycle 1 and Cycle 2. Lettered bullets are elimination criteria used or newly developed (in darker red) during that time of exploration.
pants described the design exploration process themselves gives hints to their mental models of the process, and how they choose to present design decision-making rationale post-exploration.

Two out of ten participants used an abstract visual representation when asked to explain their design exploration process (Figure 5.20). P0 represented his exploration process by a spiraling line, where the concentric circles on the spiral represent cycles of explorations. The bottom line with abstract curves was also a high-level representation of the entire exploration process that involved frequent back tracking, criteria development, and decision changes in the exploration process. The pie chart, cubic structure, and text aesthetic are detailed criteria about the selection process. The pie chart indicates a correct color balance, while the cube indicates the call for a cubism styles in the task goal.

Supporting different types of representations in design exploration processes can be useful in generative design tools to guide the designers to navigate the process visually and assist with design rationale recording. This can be especially helpful for users that already have this mental model about the exploration process.
Figure 5.21: Examples of procedural textual self description of the design exploration process: P4 and P9.

Figure 5.22: Examples of Detailed Visual Self Description of the Design Exploration Process: P1, P7 and P8.
5.3.2 Challenges in Narrowing Down Overloaded Alternatives

Eight out of ten participants reported having difficulty narrowing down alternatives. Some participants referred to the task as more focused on grouping rather than elimination. Initial criteria development was challenging due to lack of constraints and knowledge about the generated sets. Participants had to make their own assumptions of the situation and define their own rules and understanding of the alternative set.

There are difficulties in deciding about each individual alternative. Participants reported “sitting on the fence” for many of the alternatives and “questioning about my selection, [...] if it’s an error, yea, definitely an error”. Also as a concept matured, adding or removing alternatives from a group of selected criteria becomes harder due to the fear of having too many elements confusing their intent: “I want to make my elements of design as clean as possible, not too much junk, [...] my interest is in this kind of geometry, now is the new addition going to help to clarify it, or is it going to be two different kind of designs?”

Three out of ten participants mentioned wanting to go through all 1000 alternatives given to them to avoid “missing out on something interesting.” Out of the three, P2 is the only participant that finished viewing all 1000 alternatives in his session.

5.3.3 How Many Alternatives is Too Many?

One of the questions that was investigated in Study 2 (Chapter 4) was how many alternatives designers work with at different stages of design and in different contexts. We continued to ask the same questions in this study, both inquiring on generalized behaviour in participant’s regular practice, as well as more specific behaviour in this study’s context (i.e. large number of computer generated alternatives during design exploration). This data was reported quantitatively, but was analyzed qualitatively to understand at which scale the number of alternatives participants were working with, rather than a specific number or precise number range (Table 5.2 and 5.3).

Table 5.2 presents selected pre and post-questionnaire data reporting the number of alternatives participants working with in different contexts of regular practice and in this study. The great majority (except for P4) of participants claimed being able to comfortably work with under 7 alternatives simultaneously in regular practice (Pre-Q6). Most reported generating more alternatives during the conceptual design (exploration) phase than other phases during design practice (Post-Q1 and Post-Q2). Three participants, P0, P3 and P6, mentioned the exploration process as generating alternatives then narrowing them down to a fewer number of alternatives in regular practice. In the study’s context, most participants reported working with less than 7 alternatives simultaneously (Post-Q4), consistent with reports of regular practice (Pre-Q6). Some participants reported (also observed) having way more than 7 on the table during exploration while speculating alternatives, but quickly worked to select or eliminate alternatives down to a handful (7+/−2) of alternatives, or group the alternatives into a handful (7+/−2) of groups. This may be explained with Miller (1994)’s famous
Table 5.2: Number of alternatives reported by participants for different conditions.

<table>
<thead>
<tr>
<th></th>
<th>Pre-Q6</th>
<th>Post-Q1</th>
<th>Post-Q2</th>
<th>Post-Q4</th>
</tr>
</thead>
<tbody>
<tr>
<td>P0</td>
<td>2 - 7</td>
<td>100 → 3</td>
<td>5 - 6</td>
<td>N/A*</td>
</tr>
<tr>
<td>P1</td>
<td>3 - 5</td>
<td>100 - 200</td>
<td>unsure about exactly how many, but narrowing the conceptual ones down, 50% - 30% - 20% - 10% - 1 or 2%</td>
<td>1</td>
</tr>
<tr>
<td>P2</td>
<td>3 - 6</td>
<td>20 - 100</td>
<td>not 1000 at all, can generate a whole bunch, as long as I have a filtering mechanism</td>
<td>4 - 5</td>
</tr>
<tr>
<td>P3</td>
<td>1</td>
<td>4 max → 2 or 1</td>
<td>1 or 2, but based on client preference, maybe 2 or 3 that will merge into 1 very soon</td>
<td>1</td>
</tr>
<tr>
<td>P4</td>
<td>5 - 15</td>
<td>&lt; 20</td>
<td>5 to a dozen</td>
<td>12 initially, then reduced to 5 - 6</td>
</tr>
<tr>
<td>P5</td>
<td>1 - 3</td>
<td>as many as 10,000</td>
<td>as many as 10,000</td>
<td>1</td>
</tr>
<tr>
<td>P6</td>
<td>4</td>
<td>50 → 8</td>
<td>3 - 4</td>
<td>filtering: 1 only, viewing: 9 - 10</td>
</tr>
<tr>
<td>P7</td>
<td>5 - 6</td>
<td>10</td>
<td>&gt;15</td>
<td>1</td>
</tr>
<tr>
<td>P8</td>
<td>1 - 2</td>
<td>30 - 40</td>
<td>5 - 15</td>
<td>2 - 3</td>
</tr>
<tr>
<td>P9</td>
<td>2 - 3</td>
<td>2 - 5</td>
<td>10 - 20</td>
<td>layout: 20 - 25, viewing: 2 - 3</td>
</tr>
</tbody>
</table>

**Table Notes**:
- **Pre-Q6**: Number of alternatives can be handled comfortably simultaneously in design practice
- **Post-Q1**: Number of alternatives normally generated in conceptual design phase
- **Post-Q2**: Number of alternatives normally generated in other design phases
- **Post-Q4**: Number of alternatives actually handled simultaneously during the study session
- * data unavailable for P0

The Magical Number Seven Plus or Minus Two theory for the capacity for processing information, but remains as a cautious claim before further investigation.

Data from questionnaires (Post-Q5) reporting participants’ own estimates of how many alternatives were viewed in the experiment session overall and how many alternatives they focused on or looked at for more than 5 seconds. Table 5.3 suggest that the estimated viewed alternatives overall is relatively accurate while the “viewed for 5+ seconds” number is not. It might be due to the fact that most participants have a visual cue regarding how thick the unviewed stack of cards is. Also, qualitative observations suggest that participants using more implicit and non-physical recorded criteria might be under-estimating the number of alternatives viewed for more than 5 seconds, and participants using more explicit and physically-recorded criteria might be over-estimating the num-

105
Table 5.3: Participants’ estimates of number of alternatives on cards viewed (Post-Q5). ** P0 data unavailable.

<table>
<thead>
<tr>
<th>**</th>
<th>Estimated Viewed Cards</th>
<th>Estimated Cards Viewed For 5+ seconds</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>200-300</td>
<td>100</td>
</tr>
<tr>
<td>P2</td>
<td>1000</td>
<td>5 or 6</td>
</tr>
<tr>
<td>P3</td>
<td>800</td>
<td>100</td>
</tr>
<tr>
<td>P4</td>
<td>60 or 70</td>
<td>20</td>
</tr>
<tr>
<td>P5</td>
<td>120</td>
<td>17</td>
</tr>
<tr>
<td>P6</td>
<td>50</td>
<td>30</td>
</tr>
<tr>
<td>P7</td>
<td>500</td>
<td>2</td>
</tr>
<tr>
<td>P8</td>
<td>40-50</td>
<td>20-25</td>
</tr>
<tr>
<td>P9</td>
<td>100-150</td>
<td>20-30</td>
</tr>
</tbody>
</table>

ber of alternatives viewed for more than 5 seconds. More quantitative confirmations are needed after coding all participant videos.

Five out of the nine participants reported to have explored 50 to 150 alternatives (Table 5.3). This may have correlations with findings from Study 2 with typical manual design explorations using generative computational tools are in the scale of around 100 alternatives (Section 4.3.3). It is worth noting that although P3 has a reported estimation of 800 alternatives view, a great majority of which are viewed through a quick flipping action towards the end of the session.

Although still needing further studies to confirm this hypothesis quantitatively, we argue that the data obtained in this study may reveal the following for design exploration with computer generated alternatives in most CDT users:

- can only focus on one alternative at a time
- can comfortably handle 2 to 7 alternatives (or alternative groups) in parallel
- choose to explore around 50 to 100 alternatives in this context

5.3.4 Example Analysis with Interactive Visualization (with P1 data)

In a session of 1 hour 46 min and 46 sec long, P1 had viewed a total of 382 cards out of the 1000 given to him. P1 had four visible design exploration cycles shown in defined steps (Figure 5.23). As mentioned earlier, P1’s video data was coded in detailed and input into an interactive visualization built to discover patterns in his design exploration process. Figure 5.24 shows an adapted illustration from a screenshot of the interactive visualization indicating these four cycles. Each cycle is visibly getting faster than its last one.
P1 developed explicit and physically-recorded criteria (mostly on post-its, some on the white-board and task instruction sheet initially) to procedurally make his design decisions. Figure 5.25 illustrates events of his criteria development, such as creating new criteria, re-ordering criteria priority, removing criteria and combining criteria. It is evident from the graph that most of the criteria development was done in Cycle 1 and 2. In fact, closer inspection to individual criteria development event details show all criteria as finalized by the end of Cycle 2, and used throughout Cycle 3 and 4. Though P1 had reflections regarding possibly changing criteria definitions or re-ordering criteria in the last 2 cycles (seen as criteria events in Figure 5.25), P1 reconfirmed that criteria development was final and did not make any further changes. Here is a detailed description of each cycle:

**Cycle 1** P1 spent a significant amount of this time developing criteria, as represented by white blocks in Figure 5.25. He initially viewed cards one by one, slowly attempting to understand the set and the task description. He rejected some cards by putting them facing down in the right-hand corner of the table. He laid out all non-rejected cards on the table for criteria development. Throughout this process, he interrupted card viewing to record criteria down on the instruction sheet and the white board for brainstorming. Eventually he settled on writing each confirmed criteria on individual post-it notes (Figure 5.26).

When he feels he has a good idea about the criteria, he recollects all cards on the table and goes through them again using the criteria developed (Figure 5.27: Picture 1). At this time, only a few post-its were present. Most criteria were visual based on aesthetics of the buildings.
Figure 5.24: **Illustration of P1’s temporal design patterns adapted from a screenshot of the interactive visualization.** Graph clearly shows four cycles of design decision making, indicated by breaks in card movements from bottom to middle - signifying seeing unviewed cards. Sections 1a, 2a, 3a, 4a are times P1 was seeing new cards and laying them out on the table if they passed the initial criteria. Section 1b, 2b, 3b, 4b are times P1 were seeing cards on the table and migrating them down his list of criteria.

Figure 5.25: **Illustration of P1’s criteria development events adapted from scaled-down screenshot.** Stages marked on top in grey boxes are consistent with those in Figure 5.24. White blocks in the vertical center of the graph are times P1 were creating, updating, reordering his judging criteria by interacting with post-it notes.

Figure 5.26: P1 Cycle 1 - Initial speculation and strategy development, with explicit and physically recorded criteria on whiteboard, instruction sheets and post-its.
P1 then incrementally added more criteria during viewing the alternatives. He developed his filtering strategy of migrating cards from the left to the right of the table through each criteria (on post-it). All cards that failed to pass a certain criterion remained in the same column as the criterion’s post-it. Criteria were re-ordered based on how quickly it could filter out alternatives, ones that could get rid of more cards and were easy to assess were moved towards the left.

Note that subcategories created for multi-story buildings by number of floors (2, 3, 4, 5+) remained in that structure through the left to right migration. Buildings that were two-storey high were considered too short after evaluating the multi-story criteria, and left 3 different rows of cards representing 3, 4 or 5 floor buildings across the table. Towards the end of the cycle, P1 created a mini-ruler by folding one post-it note and drawing even marks on it. The mini-ruler helped him in subsequent evaluations of the exact number of floors when it was difficult to determine with a bare eye.

Picture 5 in Figure 5.27 shows the end state of Cycle 1. According to P1’s own rules at this point, he only had 1 alternative left in the last column. He decided to keep the last two columns on the table as his final selection from this cycle due to the low number that made it to the end.

**Cycle 2** P1 continued to use the same filtering strategy and criteria developed in Cycle 1. While viewing alternatives and migrating the cards from left to right, criteria were added, re-ordered, and updated during the process. P1 started from the left-most criteria (Figure 5.29). The cards on the right-most columns now were selected from Cycle 1.
Figure 5.28: P1’s “ruler” for evaluating number of floors, self-made using post-it note. Marking on the post-notes are scales 1 - 6.

Figure 5.29: P1 Cycle 2 - Followed the same strategy and criteria developed in Cycle 1, and added more criteria during the process. Left: Start of Cycle 2. Right: Nearly at the end of Cycle 2.

Detailed inspections on the interactive visualization (Figure 5.30) with the aid of criteria events (Figure 5.25) shows the filtering process sectioned by criteria. It is evident that with each criterion filter, less cards remain. Here is a summary of interesting events during Cycle 2:

- Created [2D interpretable] post-it. This was an implicit and ill-defined criterion in Cycle 1 indicted by statements like, “I can’t figure out what it is at the back”. This was then transitioned into explicit, well-defined, and physically recorded criteria.

- Created [stilts buildings] post-it. Buildings of this type were discarded previously in evaluations for [earthquake proof] criterion. After seeing a few of the building, P1 decided to create another category of stilt buildings in parallel to earthquake proofed ones to address another style that was better in possible flooding situations. It is interesting to note that P1 used a selection strategy for card 080, ignoring other elimination criteria – “Crazy stilt building! Makes it just for its stiltness! [P1 puts it all the way down to the right, skipping the last 2 criteria.]”

- Relaxed criterion threshold or makes new assumptions. It is interesting that P1 reflected upon the criteria when most cards failed it, even when his goal was to eliminate over-
loaded cards. For example, when most of the cards are failing both earth-quake and stilt criteria, P1 commented, “this is concerning...why is it most of theses do not seem to have enough...can we assume that when that connects with the floor, it actually goes into the floor?”

- Combined sub categories of [multi-story]. After realizing determining the exact number of floors in a building was not beneficial to the decision-making process, P1 combined the sub-categories into a boolean group of whether it’s multi-story or not.

- Combined criteria [real estate cost] with [# of units / bang for $$] as a versus relationship.

There was a long black-out period with no card movements towards the end of the cycle, as seen in Figure 5.29. During this time, P1 reflected on his strategy and all criteria created. He cross-referenced the instruction sheet, notes created by himself on the whiteboard and on paper repeatedly with his current impressions of the generated set. He thorough considered the order of the criteria, as well as the relationships between them. After this cycle, all
physically recorded criteria and their ordering were finalized. Same as in Cycle 1, only the last 2 columns were kept on the table as were selected for Cycle 2.

**Cycle 3** P1 continued to use the same exploration strategy and criteria developed in Cycle 2. Figure 5.31 shows Cycle 3 was visibly faster than the previous 2 cycles (indicated by shorter overall span across the x-axis), with far fewer pauses for decision making (indicated by less black gaps between vertical lines).

![Figure 5.31: P1 Design Process (Cycle 3 Detailed View), showing criteria used to filter out alternatives with the same criteria list as in 5.30.](image)

There was still a black-out reflection period at the end of the cycle like Cycle 2. Ideas of re-ordering and creating new criteria surfaced in this period. After careful considerations, P1 decided to keep criteria structure the same.

**Cycle 4** This cycle was very similar to Cycle 3, yet shorter again. It is interesting to see that so few cards got past criteria #5, and filtering for criteria #6 to #9 are done so fast that it is barely
visible on the graph (Figure 5.33). This was the last cycle of the design exploration process. Figure 5.34 shows the final state of the table at the end of the cycle that was presented by the subject. A total of 9 alternatives remained as the final options, presented in two rows - stilt buildings and regular ones and two columns - whether the last two criteria were passed or not (greenery and blockiness).

Figure 5.33: P1 Design Process (Cycle 4 Detailed View), showing criteria used by P1 to filter out alternatives. Same criteria list as in 5.30.

Figure 5.34: P1 Cycle 4 - Final state of the design exploration process.

5.4 Discussion

This study presents an exploratory observation study in investigating design exploration behaviours with 1000 computer generated alternatives. Results revealed a common cyclic pattern among par-
ticipants in their design exploration process in this study’s context. It also identified three different ways participants chose to represent their own design exploration processes. Data collected from participant questionnaires during post-task interviews revealed challenges in narrowing down alternatives during the overloaded exploration session, provided insights into the quantitative scale users’ cognitive limitations allowed them to deal with in regular practice and in this study’s context. This section will discuss the implications and limitations from these findings.

Video analysis of participants performing the design exploration task show they commonly followed the same cycle. This is consistent with design literature described in Chapter 2, that describes design exploration as cyclic (Brown, 2009) and a co-evolution process between the problem and solution space (Maher and Poon, 1996). Evidence from this study confirms that design exploration with overloaded computer generated alternatives follows the same pattern as general design exploration processes.

Qualitative analysis from video data reveals hints as to what kind of criteria are developed during what stages of exploration, and how criteria evolve over time. For the low number of participants in this experiment, no conclusive quantitative inferences can be made. However, we make the following qualitative speculative observations, that need further quantitative studies for confirmation:

- Visual criteria may be more likely to be used initially than structural ones, possibly since it is quicker to assess cognitively.
- Decisions made using implicit and ill-defined criteria may be more likely to be overturned at a later stage (Figure 5.35 and 5.36 are examples of a decision overturned during P1’s session).

![Figure 5.35: Examples of two alternatives that had been discarded initially, but were reverted back to being temporarily accepted, and eventually lasted till the end of the session.](image)

- In overloaded computer generated alternative contexts, designers may be more likely to accept solutions that they would not consider previously due to repeat exposure. Therefore it may encourage more novel and divergent solutions. e.g. P1 initially discarded stilt buildings, but started to accept them after seeing several of them. and eventually created a selection criterion for stilt buildings (Section 5.3.4).
- Designers that retain examples or grouping information for discarded solutions might generate better final solutions than designers who only retained selected solutions in the exploration
process. e.g. most participants like P1 retained one pile for all discarded solutions, while P6 retained all eliminated groups based on different criteria in separate piles and commented on referring to them as examples of “don’t” along with the selected criteria for next stages of design (Figure 5.19). e.g.2, P3 retains “do’s” and “don’ts” by different orientations of card placement (Figure 5.37). Examples of disliked alternatives are vertically placed.

Figure 5.37: Example of P3’s exploration process, in which he uses vertical placements to indicate groupings of discarded alternatives.

⋄ When too many alternatives (more than 7) are in current view (on table) in this context, designers may almost always try to reduce that number down by eliminating or grouping with existing criteria. New criteria are created at this point, if designers are unable to bring the number down using existing ones.

⋄ Perhaps in a computer generative alternatives context, the design exploration cycle terminates when designers feel criteria are finalized, rather than when designers feel no more new alternatives are coming up. The comfortable quantity range for designers to explore computer generative alternatives might be 50 to 100.

⋄ When there are no physically recorded criteria structures to guide the process, participants resort to their memory of all developed criteria till now, and are often evaluating multiple
criteria at the same time (unlike P1, who was doing mostly one criteria at a time). In this style, selected alternatives may not have passed all criteria, since cognitive limits prevents them from being able to evaluate many criteria at once. This may cause the final exploration outcome towards having more unqualified alternatives from non-visual criteria, as designers may opt for assessing easier ones first.

Despite best efforts to avoid biases, results from this study may still be compromised by having only one data coder and analyzer. Developing a more reliable coding scheme with the help of additional coders may help to achieve a more consistent coding scheme with a deeper understanding of the data. Limitations of only having P1 detailed coded will also be resolved in future analysis after the code is developed to achieve high inter-rater reliability. There might also have been artifacts introduced by the procedures. For example, the required overhead GoPro camera worn by participants may or may not have influenced their design exploration behaviours for possible factors like self-consciousness, physical discomfort from wearing the headband etc. All models, sample “inspiration” images, task descriptions and conversation opportunities with the researchers during the task can all introduce artifacts in the final results. Although I have not discovered any particular concerning artifacts to date in the listed concerns, it is still possible that there are unknown effects that may limit the results further.

5.5 Conclusion

In this study, we observed how designers performed an architectural design exploration task with a large number of computer generated alternatives from a parametric model. We attempted to describe the cyclic behaviour pattern based on data from all participants. We also try to identify differences in their criteria development process and self-descriptions of the process. This study also includes efforts in developing two interactive visualization tools to assist the researcher in designing the study and analyzing the data. The first one is a mix-initiative visualization that assist the researchers to explore different methods for generating a set of 1000 alternatives with great visual diversity from a parametric model (described in Appendix C in detail). The second one is an experimental interactive data analysis tool for interpreting video coding data for similar design exploration studies. Findings in design cognition and tool building from this study provide insights into developing future computational design tools (to be discussed further in Chapter 8).
Chapter 6

Overloaded Design Exploration in CDTs

In this thesis, I have investigated how designers interact with design alternatives during design exploration. The last two studies, especially Study 3, focused on how designers explore overloaded number of CDT generated alternatives (i.e. overloaded design exploration or ODE thereon). This chapter will present my take on the implications and contributions of these findings.

I will start by going back to the Schön and Simon discussion in the literature review (Chapter 2). Does this thesis provide evidences for or against each of their views? I think yes, we can see evidences supporting both. This also echoes with my interpretation of the two viewpoints from Schön and Simon, they are actually non-contradictory explanations of design exploration behaviours from different angles. Study 3 showed a common exploration cycle with both reflection-on-action and reflection-in-action patterns in their criteria development process (Schön, 1983). For example, in P1’s detailed analysis, we find new criteria being created or old criteria’s ranges being relaxed in-action while P1 uses these criteria to select and eliminate alternatives. Each design exploration cycle ends when P1 runs out of cards to evaluate for this round. Observations showed P1 reconsidered all current criteria at the end of every cycle, to review on-action for each criterion’s priority, validity and relationships with other criteria. This observed cycle is also in line with Simon (1996)’s generate-test cycle in design processes, although the generator in this case is the CDT, while the participants take on one or more testing cycles for evolving constraints.

The next question is whether overloaded design exploration (ODE) is any different than non-overloaded design explorations (NDE). When designers are presented with overloaded number of alternatives, the cognitive resource limitation becomes more apparent than in non-overloaded scenarios. Simon (1996)’s’ model of design decision making states that human cognitive limits restrict designers to find satisficing solutions, rather than the most optimized ones. We found that in many situations in Study 2 and Study 3 (overloaded), participants are only able to attend to a subset of the criteria developed during selection or elimination processes, while in Study 1, (non-overloaded) participants are more likely to consider all criteria on all alternatives.

Another speculation is that designers’ perceptions differ between ODE and NDE when involving alternatives generated by others. In NDE, designers perceive the exploration process of alternatives
generated by others as an evaluation process, just as the alternatives were generated by themselves. However, in ODE, it seems that designers are using the volume of alternatives as a knowledge base to understand the design problem. Many participants in Study 3 commented that they want to see more alternatives to figure out what the design task is asking them to do. There is a perhaps unconscious assumption here, that whoever generated the alternatives must know what they are doing, the alternatives at hand in the ODE must contain satisfactory solutions. I cannot generalize this speculation with certainty until further experiments confirm the same finding, as it may be an artifact from lab participants trusting the researchers. Nonetheless, I think this is an interesting aspect of ODE that might be different from NDE.

Furthermore, reported by participants from Study 2 and observed in Study 3: in ODE cases, designers tend to establish criteria or an example set for both “do’s” and “don’ts”, while in NDE cases, designers tend to focus only on “do’s”. Also, there seems to be a tendency of sorting alternatives into piles in ODE than NDE. I believe this is a natural chunking strategy to overcome cognitive limits of dealing with overloaded number of alternatives. These are also in line with Simon (2000)’s discussion around limits of adaptation to task environment:

“A bridge, under its usual conditions of service, behaves simply as a relatively smooth level surface on which vehicles can move. Only when it has been overloaded do we learn the physical properties of the materials from which it is built.” (Simon, 2000, page 13)

Last but not least, **how to apply these findings to CDT design to support ODE?** I attempt to make some design recommendations. Keeping Blackwell (2002)’s attention investment model in mind, I would like to suggest some high-level interaction abstractions that have low costs for the users to perform, yet with immediate increased benefits for design exploration:

- Supporting *reflection-in-action* and *reflection-on-action* to encourage conversations between the designers and the alternatives in the generate-test cycles. For example, we observed that some participants used a sequential criteria development pattern. They filtered alternatives with different criteria one by one to select the final set of alternatives. Figure 6.1 illustrates how to support the progress by organizing the stream of alternatives while they going through the cycle.

- Providing interaction models for designers to explicitly record developed criteria and his or her state in the exploration cycle. For example, users may be allowed to annotate criteria in freeform (e.g. natural language or a simple sketch), rather than being forced to annotate it in computational terms of parameters and conditionals.

- Identifying which criteria are used more often in order to determine if the alternatives are *satisficing*, and monitoring possible biases resulted from limited cognitive resources to perform full *optimization* (e.g. using aesthetics more than structural). For example, when criteria are all explicitly recorded, it would be easy to see which criteria are used more often than others.
in the process. During the reflection-on-action phase, the designer may realize an overuse of aesthetics if he wishes to correct that bias.

○ Implementing features for grouping alternatives by user-defined criteria. For example, in Figure 6.2, I show the possibility of representing criteria as sets of alternatives that are mutually exclusive. This will help users in the scenarios where they want to have groups of alternatives classified by their traits for different criteria. For example, when a user tracks “do’s” and “don’ts” as different alternative categories, he can put all alternatives into two criteria sets that represent the two sets. There is also a possibility of splitting and merging sets in this representation.

Criteria as a “set” of alternatives

○ Supporting designers to use alternatives as a way of figuring out what the design problem is. When a designer is overloaded with a large number of alternatives that are not generated by himself, it appears that he will initially use the design exploration process as a way of understanding the design problem rather than seeking a solution. During this stage, it is important for the designer to see the relationship between different criteria, and to sort alternatives within each criteria. Figure 6.3 and 6.4 show two different interaction styles that will support the designer to understand the relationship between alternatives and criteria.
Figure 6.3: Re-arrangement of alternatives’ ranking within each criteria during design exploration.

Figure 6.4: Criteria used as a "tagging" attribute. When an alternative is dragged and dropped from one criteria to another, it is "tagged" with both criteria, instead of leaving the first one. This way the alternative can belong to different criteria simultaneously.
While these are interesting findings that still need more confirmatory studies to establish its grounds, we must also recognize that new knowledge are being created every day to overturn our understanding of how information is processed. For example, recent publication Yamagata et al. (2015) overturned what we believed about short-term memory (STM or working memory) and long-term memory (LTM) for decades. It was long thought that LTM are created from repeated exposure and recalls, but Yamagata et al. (2015) found in a fruit fly study that LTM can be created with a momentous experience, without repeated exposure and recall. If this mechanism is proven to be repeatable in humans, the entire overloaded design exploration process may need to be revisited. Perhaps criteria that are capable of creating LTM with single exposure are used more as the *satisficing* criteria. To stretch this further, if LTM can be created without going through working memory, then with the correct triggers, humans will not be bound with limited working memory as prominently recognized in Miller (1994)’s magic number. In that case, design exploration may need a larger threshold to become overloaded. Basic cognitive science research that understands the details in memory and perception is strongly called for to investigate the mechanisms of overloaded design exploration.

Here I have discussed highlighted findings of overloaded design exploration. In the next chapter, I will discuss more general findings and UX suggestions for CDTs that can be applied for both overloaded and non-overloaded contexts.
Chapter 7

Lessons Learned For Designing Computational Design Tools (CDTs)

The journey I took in writing this Master’s thesis was one full of learning all around: from methodologies, to different design disciplines and design processes. Now at the end of this journey, I would like to share my thoughts on how to design better CDTs based on lessons I learnt through literature review and conducting these studies. Many of these are speculations based on “leap of faith” to the next step of the scientific investigation. Some of these lessons may even be known principles to experts in the field, but new to me in this journey. Nonetheless, I would like to share these six “design recommendations” for future CDTs.

7.1 Support Modular Alternative Explorations

Interviewees from Study 2 (Chapter 4) reported that design exploration could happen during all stages of design, and often for a specific aspect of the design problem at hand. During exploration, most variables of the design were fixed while the designer manipulated a few variables to create new alternatives. I term this process as modular alternative exploration. Currently, designers accomplish this type of exploration by either 1) duplicating the design artifact for one alternative multiple times, and revising each one manually to reflect different alternatives, or 2) manually changing parameter values to generate different options. Neither of these interactions are easy for modular alternative explorations. Not only being a labour intensive process, the communication and recording of such alternatives are also difficult. Especially in larger project, different team members working on generating alternatives of various parts or levels need extra efforts to synchronize the alternatives. I believe future CDTs should lower the barrier of both approaches of modular alternative explorations.

I suggest that CDTs should provide users an easy way to “fix” most of the project while creating alternatives for a specific aspect or part of the project, i.e. modular alternative exploration as defined earlier. This requires the tool to know the state of each component as being a constant or
variable during each exploration. The states recorded as part of an on-going project history will create a version-controlled project repository mechanism for parallel working branches of alternatives (Figure 7.1).

Figure 7.1: Example of two parallel modular alternative explorations (i.e. windows, exterior colours) and how they can be represented and merged with a version-control structure.

### 7.2 Recognize and Support Different Roles in Design Team

From the interviews (Study 2, Chapter 4), I identified multiple user roles in the design process. Each participant, especially for more complex projects, may have one or more roles on a dynamic design team while using Computational Design Tools (CDTs). Due to the small sample of this
interview study, I think there may be more roles that are not identified. However, even with limited findings, I believe there are enough evidence that current design tools should develop into more specialized roles to serve each distinct user role to its best, while maintain an integrated form to work seamlessly with each other. Existing approaches seem to either lean towards developing a “good-for-all” product has an average experience for everybody, or being too specialized that only a group of expert users benefit from it. For example, AutoCAD suite of software has been reported to be taking the first approach, and Grasshopper is taking the second. It is a challenging task, but I think these systems should be supporting more diverse user profiles. Even without developing more functionalities specific user roles, these general systems can benefit from developing special user workspace profiles and interaction flows.

The minimum number of user roles I believe should be supported are:

- **Sketcher**: Focus on producing initial sketches or sketching modifications to the existing artifacts. The word “sketch” here is referring to the action of translating rough ideas onto an external medium. Sketching can be a traditional approach with pencils on paper, or on digital interfaces such like SketchUp.

- **Drafter**: Focus on transferring a sketch created by the sketcher. The sketcher and the drafter can be the same person or different members on a design team. However, the activities required in the drafter role is directed towards more formal and detailed representations than the sketchers. Some of the functions from sketcher or drafter may overlap, especially when the sketching is also done in the same digital interface.

- **Generative model designer**: Focus on designing a generative model, rather than a particular design iteration. Higher functionalities of programming relationships between parameters, manipulation values for quick previews of models is emphasized more than concrete drawing functionalities, such as rendering. Also the tools need to provide easy interactions with high-level properties of the model. Specifically in version-control to view comparisons of revisions, should be possible. In some cases (especially for consulting firms and R&D support teams in larger firms), parametric models are built to generalize a type of problem that can be reused in different projects, however, the parameter model designer aims to build a custom piece of software that will work in a spreadsheet fashion for others using it. Here, often using parametric model to explore alternatives may not be the goal, rather, it needs to produce a single robust solution. In these instances, the problem can very well be solved through other manual methods, the parametric model is only built or revised for future cases of similar projects.

- **Generative model value iterator**: Focus on creating design alternatives by iterating values in an existing parametric model. System features of this profile could be used in conjunction with the other profiles for previewing generated solutions while the designers is experimenting or “debugging” the model. Different modes of value iterations could include: 1)
changing value one at a time, much like the trial-and-error approach, 2) generating a set of pre-determined values (e.g. incremental within a certain range for each parameter) at the same time, or 3) a combination of the two by switching from one to the other. Regardless of which mode of value iterations is used, this user profile must be strong in side-by-side comparison of one or more alternatives. A much needed functionality is to provide an overview to all generated alternatives, particularly when if the number gets too large to be meaningfully displayed on a limited screen space. Supporting navigation through a set of generated solutions is important.

◊ **Viewer:** Focus on viewing results from sketcher, drafter, model designer or value iterator, rather than directly manipulating the creations. This user profile is mostly targeted for presentation purposes to other stakeholders, such as clients or decision-makers in a collaborative project. Functionalities of commenting and annotation is important. Permissions to view or edit should be set by the presentation maker. When edit permission is allowed, the viewer profile would be easily switched to other profiles to realize modification requests in comments.

### 7.3 Support Different Numbers of Alternatives

Findings from the three studies (Chapter 3 to 5) show that designer work with significantly different number of alternatives, i.e. from having only one solution in some cases to having thousands in some. CDTs should recognize that at different number scales of alternatives generation and evaluation, user goals and needs have different focuses. Visual and interaction design of CDTs need to cater to various needs, yet remain accessible during mode switches.

CDTs should support different user needs at different number scales for displaying and interacting with alternatives. For example, interviewees in Study 2 (Chapter 4) reported that single solution scenarios can occur when project resources are limited. A major user need in single solution cases is to convince the client to agree with this only option with as little revisions as possible. On the other hand, when dealing with less than five alternatives (the most common case reported), the needs shifted to comparisons between the alternatives in detail, and reflecting on whether more alternatives are needed. CDTs should recognize the difference focus: presentation and rationale recording (both real-time and retrospectively) for single solution, alternatives side-by-side comparisons, reflection and evaluation for less than 5 alternatives. Also in Study 1 (Chapter 3), all participants fall into these two scales (i.e. single solution, and less than five alternatives). These will require different visual layout and interaction flows of CDTs.

For larger numbers of alternative solutions, around or less than 100, the focus tends to be manual selection and filtering from the set to reduce alternative set to a manageable size. Although with the same goals for alternative comparisons and evaluation in detail as the less than five alternatives scenarios, there is an additional need of scanning through alternatives quickly, and seeing alterna-
tive clusters in meaningful visual arrangements. As reported by the interviewees, during the manual selection, they are interested in exploring possibilities to show the client, hence are more interested in identifying extremes in the range rather than finding one ideal solution. plausible solutions. For even larger scales of alternative exploration, 200+ commonly hundreds or thousands, the goal of the CDT user changes again. Interviewees report that at this scale, typically, they are looking for a systematic and algorithmic way of reducing the set. Identifying relationships between the alternatives and their programmable parameters in CDT is crucial in moving towards devising automated filtering to reduce the set size.

Future CDTs should consider the differences in user experience requirements for working at these different scales of number of alternatives by supporting each scale as described.

7.4 Encourage Integration of Software Practices into Design Practices

Software practices have moved towards a more agile and open-sourced environment in the last two decades. I believe that design practices can benefit from moving the same direction. Some design disciplines, such as web design, and social design (e.g. crowd design on OpenIDEO for solving social problems) are already benefiting from this. However, more traditional disciplines such as architectural seem to have not fully progressed into more agile and open-sourced ideologies.

One of the interviewees from Study 2(Chapter 4) proposed for the architectural industry to follow the open source movement as the software industry:

“I think in computer science, there’s a mindset of sharing and supporting. You never go write a library from scratch, people just write it and share it you know. but in the building industry, this mindset is not there yet. in the field, they think the differentiator are still that I have something that nobody else knew about it yet, which kind of doesn’t help with us to do better buildings. First of all, these are not locked together. Everybody has their own small things, that are not the all the picture, and they never come out and be ready to share everything, to see how it can fit together. It kind of stops us, the whole building industry from going up, that’s what other industries are experiencing now by sharing all the stuff. You know I’m a big fan of open source projects. so this is the main difference between how Linux and Microsoft works, from Microsoft point of view, we can’t hire someone that’s only going to work on something for only 2 months. I should have you to work on something for 1 or 2 years, because I hired you, right? But in an open source environment, like Linux, if you’ve done something, even something crazy like the kernel, you just go to the back and share it, you are done. that’s why systems like Linux are crazy good, because there are a lot of people who are good at something and just do that for you. If you can’t see that, in the building industry, that’s too bad for you. Different firms are working on different things, which are really good at some parts, but they don’t know about the other parts, and that is normal. If we start to see all the parts, at some point starts sharing their knowledge in a way that can help others
to make the bigger infrastructure better, then we made it to the next level. It’s a market, we want to make money. But again, the only way of making money, at least if you ask me, there are many ways, but to me, the way of making money is not keeping stuff secret, but more about sharing. ” - P4 making a case for open-source design.

Interviewees from Study 2 (Chapter 4) also report client communications and satisfactions are one of biggest challenges in design. I venture boldly for suggesting to build CDTs with more client-testing concepts in mind. Although the current architectural design practice involves a lot of client inputs and communication cycles, it is not nearly as progressive as software industry user-testing practices. Some may argue that the two industries are very different, especially in the type of artifacts that is being developed. However, I argue that architecture and software design is much the same, in that the end goal is to serve the clients to the best with the input of professional designers. Currently consultations with the clients are only happening at key milestones of the projects, with the assumption that the client would be mostly pleased with the solutions given, only small modifications are needed. Designers report that they often find themselves defending their proposals to convince the clients, anticipating more agreement than modifications. Although the end result is the same for the clients to be agreeing with the design deliverables, in practice this experience could be improved by allowing more client participation in between iterations.

7.5 Enable Effortless Transition Between Digital and Physical Artifacts

All three studies suggest that designers interact with physical artifacts at one of more stages of design. Incremental design modifications happen while the artifacts are in physical or digital, or during the transition in between. I believe it is essential for design environments to incorporate more system features to import and export information from physical artifacts where alternatives are created and tested in multiple forms. Some obvious examples are to make it easier to print 2D or 3D artifacts from digital models, and support artifact conversion to digital models through 3D scanners. Also with recent advances of low-cost virtual reality technologies, such as Oculus Rift and Microsoft HoloLens, the 3D pseudo-physical artifacts rendering may become more prevalent to blurring the traditional digital-physical boundaries. These are digital artifacts that enable physical interactions like true physical artifacts. Some other examples of improvements in enabling effortless transition between digital and physical include:

- new interfaces to easily incorporate sketches made on a CAD print back onto the digital CAD model which capture history and variations
- support more open formats that are better at the task of transitioning between digital and physical artifacts, through which alternative solutions are branched in the process for revealing opportunities

127
In Study 2 (Chapter 4, participants report that some CDTs have an unrealistic user model of keeping the entire design interaction process within the digital realm, therefore have very poor transitioning functionalities. Tool proficiency is crucial for that level of integration.

Furthermore, even when all designers become proficient with digital tools, physical artifacts are still needed because they provide different affordances that cannot be replaced by digital ones. Major revisions to existing artifacts may happen during a conversation involving two or more design stakeholders. This is reported as happening frequently in team meetings by participants in Study 2 (Chapter 4) as well as observed in Study 1 (Chapter 3) and Study 3 (Chapter 5) when participants present their final solutions. CDTs should support scenarios where one person tries to draw on existing print outs or sketches of a design to illustrate a new idea that was generated before or during the meeting. Others may join in to continue sketching on the artifact during verbal exchanges of ideas, to further clarify ideas proposed earlier or to add thoughts of their own. CDTs developer should consider user interactions with all types of artifacts, and make decisions on whether the CDT should support such interactions, or record and transition these interactions into the CDT environment.

We recommend CDT developers to recognize the boundaries between physical, VR, and digital artifacts, and provide transitions in and out of the tool smoothly (Figure 7.2). Using different forms of artifacts in design exploration may incubate more design “aha” moments, which may not be possible when restricting to only one form of representation.

Figure 7.2: CDT’s should support easy transition between artifacts during any phase of the design, including design exploration.
7.6 Build Trust Between Human and Machine.  
Do What Each Other is Best At, Together.

Psychology experiments (Kahneman, 2012, Chapter 21) consistently shown that we, human as a race, can be over confident about our skills in certain tasks even if we are fully aware of statistical proof of machines out performing us in the same tasks. For example, doctors are shown statistics of their own performance is worse than a computer program’s performance on diagnosing a certain disease. Then, the same doctors are given the computer tool to assist with their diagnose. Even knowing the tool is better than their own performance, in practice, they still manage to perform worse than the tool because they manage to out-talk themselves using the tool when faced with a real patient. They think to themselves, “but this is a different case, there is no way the computer program is going to capture this complexity, I am an expert, I think I’ve got this one the computer can’t catch”. This mentality is rooted from distrust for computer programs. Similar experiments are done with financial advisors. The interview study shows that in design disciplines, this distrusting mentality is more wide-spread and more intense than other more mathematically oriented disciplines such as financial advising and engineering.

For CDTs to work at all, initiating enough trust for users to start trying is a difficult first step. Maintaining that trust is even more challenging. Users are more tolerant in human errors (especially those from themselves), and trust that errors will be fixed next time. However, once a computer program makes a mistake in one aspect, it is assumed that the error will be repeatable. Especially in larger more complex systems, where users do not understand the mechanisms of the program, users often are ready to jump to the conclusion that the program will not only error in the same aspect, but in all aspects, and abandons use all together. For example, if on a calculator, 2+4 outputs 2 as the answer, due to a faulty 4 button, you will be ready to chuck it out for any calculations without knowing it’s 4’s fault, even if 2+8 is correct.

7.6.1 Constant reinforcements of performance merits to users

P3 stated a technique that he uses in establishing user trust and acceptance towards generative design technology, much like Kahneman (2012)’s experiments. P3 would first ask the architects to make their first attempts, provide initial sketches before using the generative tool based on their expert knowledge, e.g in energy analysis. Then P3 would use the generative tool to produce a set of best and worst solutions, and fit architects’ initial sketches on the same graph. It was found that architects are generally a lot worse than they expected in coming up with energy efficient designs. When shown performance plots including both architects generated and computer generated alternatives, architects become more receptive to the idea to using CDTs to generate alternatives and also realize the potential of having more diverse and higher performance alternatives.
7.6.2 Make procedures more transparent

One of the many reasons that “Clippy” from Microsoft Office failed was that it was not transparent enough about how it made its behavior decisions. It failed a couple of times at different tasks, people assume it’s good for nothing. But we might’ve been all mistaken, it was probably very good at some tasks, we just never found out. “Google Now” as service that was developed also as an intelligent personal assistant on the mobile phones makes the decision process more transparent. It tries to lists why it suggests these information “cards” (e.g. links, stories, game scores, weather, directions etc.) to you. For example, “these are updates to websites you recently visited”, “this is a maybe appointment from your emails”, “these are popular with readers in your area” etc. If the user does not like one suggestion, he can tell “Google Now” about it. It will ask the user if it is due to lack of interest in this particular site, or this entire category - lack of interest in getting updates from recent websites. This services builds a lot more trust with the users than “Clippy” from decades ago, because it is more transparent and it listens. Many generative CDT’s are suffering from the problem from “Clippy”, in that the technical procedures are too difficult for users to understand or customize to fit their needs. Users stop using complex functions when things mysteriously happen (or not) for reasons not transparent to them.

7.6.3 Involve user decisions as much as possible

Studies have shown that a sense of involvement from the user can add trust to the final decision, hence to the entire system. Computational design tools should have more user decision points whenever possible. Interesting thought for further studies is to test how falsely-believed sense of involvement works in these cases, much like a placebo effect. In order words, if the system made the person thought they had a conscious input into the decision process, but actually in the background runs its own decision mechanism, will the designers have more trust in the final decision. Of course, building such deception mechanisms into a system is ethical or not is another question.

7.6.4 Do things the users way

There are certain user habits that have built over the years, even if a new approach is proven to be better rationally, it may take some candy-coating to get the users to accept it. For example, Google came up with Project Fi, a revolutionary wireless and cellular service carrier plan (like AT&T, Rogers, T-Mobile or Telus). Consumer research reveal that average customers usually pay for a data plan that is well-above their actual usage. For example, pay for a 3GB plan, but only uses 1.4GB per month. Project Fi tries to introduce a way to reduce people’s cellphone bills by only paying for actual data usage. Sounds really cool, but very poorly received by users in their initial studies. Why does the user want to pay for an expensive 3GB plan that he’s not using, but not willing to pay for only the actual usage of 1.4GB data? Surprisingly, it’s because users like to think of cellphone bills as a fixed monthly cost. Also the ridiculous overage fees from traditional carriers have trained the users so well in accepting paying extra for data they don’t use, just so that
they don’t have to worry about going over the monthly limit when they are watching online videos. So Google dressed up their paying model to fit the users need. Now they offer the user to pay the same fixed bill as traditional carriers (e.g. for 3GB, or whatever they are comfortable with before). The difference is, at the end each month, Google credits back the unused data amount (e.g. if 1.4GB used, then 1.6GB will be refunded) in cash to the user’s account for his next bill. It’s like magic - Google still charges the same amount of money for the same amount of services, users don’t like it before, but do now.

Some of the CDT features are suffering from the same problem in getting user trust. The mental models of design processes have established outside of the digital tool realm, and when CDTs, especially generative ones, attempt to offer a new model, it needs to be very careful about its presentation. For example, in the interviews, a common concern with parametric tools is its difference with incremental design development outside of the generative parametric environment. Traditionally, the incremental alternative generation process happens when users can see it, while with parametric tools there is an uncertainty of outcomes during the development of the parametric model, which usually takes longer to develop than manually designed alternatives. There is always a sense of insecurity, whether it’s going to have a bug or not, how it is going to turn out, when the boss or the client calls one potentially has nothing but lines of code, and the fact that it is still running. “In a normal setting, for design, you always have something. SOMETHING, at the end of the day. you can’t really say, we have something running, we will have something in 10 days, if everything goes fine.” A direction to “dress-up” this aspect of CDT is to perhaps to provide more assistance in giving pointers to how the result will look like during the process of building the model. This is a complex problem that might involve:

- visualizing outcomes of one or more parameter ranges on the model
- creating parametric development environment with more tolerance for uncertainty, e.g. like how “code stubs” are allowed in programming environments
- more open source helper libraries and established “abstract class definitions” e.g. Java standard libraries

For example, if one is to build a parametric model for a skyscraper, the system should make libraries available to quickly construct the structure, building skin etc; and build a basic model in minutes, and then define or replace each section with customized code. Hence, if the designer decides to explore alternatives for a unique design. There will always be something to show the boss and clients without having completing the entire design. This mimics the rough sketching than detailed sketching methods where designers always have something to show and incrementally works towards the end solution.
7.6.5 Know what the users are the better at machines.

This is a very general recommendation for CDTs. I urge researchers to think about this when building CDTs and how to combine the talents to achieve great designs. While I mentioned CDTs are much better at human than estimating which building designs are more energy efficient, there are a lot more tasks that CDTs should build to tailor the interactions to let designers do their job and transition the artifact back to the system. An example from the interviews surfaced is renovation projects of older buildings. One subject commented that it is almost impossible to do these projects using computational design tools, because these original drawings are not done in CAD, and are so old that sometimes are not accurate anymore due to undocumented renovations, wear and tear etc. If CDTs were used, it would be very time consuming to first establish the current building plan, then revise it for the new renovation. Many of these projects are done quickly by designers with minimal CDT use. Hence, if a computational design tool tries to include a generative tool for these type of renovations, it would probably have to ponder on how it would help the designer. Probably not much other than assisting with photo reconstruction technologies to help the designers to document the site better, so there is more information to work with back at the office.
Chapter 8

Conclusion

Design exploration is a complex process that has characteristics across different disciplines. This thesis attempts to understand these processes with a focus on styles of alternative generation and Computational Design Tool (CDT) use. Through two exploratory observational studies and one interview study, I hope to identify how designers behave in design exploration processes, and how to build better CDT tools to better support their processes. The outcomes of this study may contribute to the growing knowledge base of describing design exploration. There is an emphasis in the UX guidelines to consider CDT users’ capabilities and needs.

Each study has its own limitations as discussed in its respective chapter. Many interesting (some speculative) observations need further quantitative studies for confirmation. However, they provide evidence for existing variances in design exploration processes, in terms of:

- styles of alternative generation
- number scales of alternatives being generated
- CDT user roles
- criteria development and design decision-making processes

The classifications of CDT use and criteria development form basis for a framework for discussing and conducting research in design exploration. They are certainly not exhaustive, but have provided some insights that to my best knowledge are not explicitly documented in existing literature. The UX guidelines proposed are based on both existing literature and evidence from the three studies presented here.

The interactive visualizations built in Study 3 and transcribing tool built in Study 2 were intended to assist researchers in devising the task, and data analysis. These can also help similar design studies in the future. More importantly, lessons learnt from interactive visualizations in Study 3 can be transferred into CDTs to provide visual overviews of design flow. We see potential in these visualizations for strategy development in design exploration, especially when given a large number of CDT-generated alternatives.

Future studies are also discussed in each individual chapter. A general direction to carry this work forward is to build prototypes of CDTs adapting the proposed guidelines, and continue with
the identify-propose-build cycle (as seen in Figure 1.1). Furthermore, detailed investigations into the variances in design exploration processes listed above are also interesting. For example, when and how are each style of alternative generation or criteria development strategies used.

Another line of research should investigate correlations between design outcomes and different design processes. For example, when given a large number of CDT-generated alternatives, are designers more likely to accept solutions that they would not normally consider due to over exposure to that type of design? Is it encouraging more novel and divergent final designs? How to control this priming effect caused by CDT on designers’ final decisions?

Overall, this thesis has been a tremendous learning experience for me in designing and conducting scientific research using both quantitative and qualitative methods. I hope findings from this study could be helpful for future studies in design exploration and inspire better CDTs.


Emel, G. (2015). Floor plans 2nd floor. https://www.flickr.com/photos/hillsdalehouse/4236794/in/photolist-nHsb-pxsp1u-8xfEzb-fYd77J-dBtmiL-c2yz33-7bCzmp-dPmqZz-gyb4HF-bnrNTQ-m2u7T1-gjhy6J-9TPv9m-9TQwmt-oTd9Fw-pZosop-77dyZk-bJT8og-9PyE3-oTxqS-8HRfhi-c7Emdk-4VoTS8-rnLMnx-8xaUwD-9y3qs-7zbD2-7gpJUE-6M5C3M-6B4Gmg-dcxXr-h7Dfjr-hBFtdb-h7FQCH-h7DgUT-h7EyCS-h7EG1q-a615gC-64V8mp-rThCQq-h7DAeh-h7EBfg-h7FQq1-h7ENKs-h7FSj2-h7EmWz-h7Fmav-7VykYq-7gkPnPn-6n7Y4u/ [Retrieved on July 22, 2015].


Gorimon (2011). 110803_053. https://www.flickr.com/photos/gorimon/6027188960/in/photolist-abAVi3-abAZBC-92jHqj-abClu3-abC1uo-mNWXUz-mNX8nV-7adqUb-abyu92-abytMn-abyGb6-abBMMW-4QsHsj-7Vkc3C-9CyWDG-jNFm7-abyYvn-abyFcz-9BxkrP-abdM6q-az3gr-abB8uq-abyuAr-ab24tP-abz2tn-ab5scf-abz6q6-abBVAL-cvkga5-ciPoRE-abYUF-abyV2-5KpshM-abzarB-abz69F-6ZDMoc-q8Vwdr-abBqwu-abBzio-abz9AP-abAUT1-oxQzd8-abBfpS-ab5a6-abAT7y-abAU3j-iKVFDn-5w9ykv-dsySHV-czPBej [Retrieved on July 22, 2015].


Larsen, M. (2013). (Im)possibilities of courageous creativity in comparative and international education research. 42(1).


Appendix A

Study 1: Post-Task Questionnaire

The following 2 page post-task questionnaire was conducted in Study 1. Please see Chapter 3 for details.
Post – Task Design Style Questionnaire

Subject #: _______________________   Date: _____________________________
Location: _______________________   Time: ____________________________

For each of the questions, select one that best describes your behaviour or thinking process during the design task.

1. **When I first started the task, I**
   a. felt lost, and did not know where to start
   b. recalled about similar websites I had seen / worked before, and followed their patterns
   c. immediately followed a routine design pattern that I normally follow
   d. thought about what is the most important aspects to be addressed
   e. followed my intuition

2. **Please rate how you feel about the level of difficulty of the task.**
   
   1   2   3   4   5   6   7
   Easy   Difficult

3. **During the process,**
   a. I had just one idea, and kept on modifying it
   b. I had multiple ideas, but only sketched one down initially, and kept on modifying it
   c. I drew multiple ideas down, and worked on them in parallel, until I decided on one as my final solution
   d. I drew multiple ideas down, and worked on each sequentially
   e. I wanted to do one of the above, but I ran out of time (please specify which one)

   If you had more than one idea, specify number of ideas you had in your mind but didn’t sketch: ____________ & number of ideas you sketched on paper: ____________

4. **I am confident that my design meets most of the functional requirements of the design task.** (Please rate your level of agreement to the statement by circling one of the following numbers)
   
   1   2   3   4   5   6   7
   Agree   Neutral   Disagree
5. Based on the descriptions of “Dreyfus Model of Skills Acquisition Stages” (following page), which one do you think you fall under in the domain of the given task:

1 – Novice
2 – Between Novice & Advanced Beginner
3 – Advanced Beginner
4 – Between Advanced Beginner & Competent
5 – Competent
6 – Between Competent & Proficient
7 – Proficient
8 – Between Proficient & Expert
9 – Expert

6. I believe users will enjoy my design very much (Please rate your level of agreement to the statement by circling one of the following numbers)

1  2  3  4  5  6  7
Agree Neutral Disagree

7. Please rate how you feel about the novelty of your final solution.

1  2  3  4  5  6  7
Very Conventional 3  4  5  6  7 Very Novel
Post – Task Background Questionnaire

Subject #: _______________________ Date: _____________________________
Location: _______________________ Time: ____________________________

1. Gender: M F

2. Age: __________

3. Current Occupation: ____________________________________________

4. Use of Computer: _________ hours / week

5. Time Spent Browsing Websites: ____________ hours / week

6. Experience with Web Page Design: Y N

   If yes, specify experience based on time:
   _________________ hours from __________ non-paid projects (school
   projects, helping friends to set up a website etc.)
   and ________________ hours from ___________ paid projects
   (professional websites, for work etc.)

7. Experience with Sketching: ________ (choose one of the following)
   a. Little/Minimum
   b. At least 2-3 years of Fine Arts Training/Equivalent (including
      highschool)
   c. Formal Fine Arts / Design Training / Equivalent Beyond highschool
   d. Fine Arts / Design Professional
Appendix B

Semi-Auto Transcription Tool

I built a mixed-initiative, semi-auto transcription tool (Figure B.1) with the intent to speed up the transcription process. I will briefly share lessons I learnt via building and using this tool, for others who might want to try this semi-auto transcription approach for similar studies.

No formal evaluation was done on if there was significant speed improvement from the tool. The subjective evaluation from the coder (also the interviewer) is that this approach is helpful with American English speakers, which seems to be what Google Web Speech API was used for. However, in this interview study, there was only one such subject, so mostly it was not significantly faster than transcribing manually. Also, initially it was under looked that though Google Web Speech API was fairly accurate in its native context (i.e. short voice instructions for Google Services, such as “OK. Google. Navigate to the closest gas station”). I think perhaps more complex sentences like transcribing interviews may not be its optimized use case, since videos must be broken up into to smaller pieces for accurate transcription. The interface (Figure B.1) was designed for operations with buttons or with keyboard shortcuts alone for:

- Copying current timestamp to transcript file, play and pause, skip back or forward for \( x \) seconds in the video (textboxes in zoomed-in view)
- \([\text{breaksentence}]\): pause auto-transcribing and allow user to edit generated script in the textbox
- \([\text{push, startnext}]\): after editing the auto-transcribed sentence, push the correct version to a temporarily paragraph

While not being a professional transcriber, and negligent about any better alternative transcribing methods, I decide to share this approach with other researchers. I suggest other researchers to give this semi-auto transcription method a try when most interviewees are American English speakers, for possible time improvement, or as verification against human-transcribed text. Two unexpected advantage I find using the custom-built tool for either American or Non-American English speakers were:

- Auto-correction of word within context that the coder had no idea about. Example, the word “photogrammetry” was used by one of the participant one or two times. The speed of conversation was fast, and the speaker was rather quiet and had an unfamiliar accent to the coder. Not
knowing the word, the coder would confidently transcribed “photo geometry”. But Google Web Speech API correctly identified the word. The coder gained better understanding of the context.

- The need to break sentences into smaller pieces for the API forced the coder to stop more often than during a manual transcription session. This forced stop of listening to shorter sentences each time reduced the amount of rewinding during normal transcribing sessions. Also, even if the auto-transcribing API could not produce a perfect transcription during one particular sentence, the phonetic representation of the rendered text is usually close enough to the accurate transcription that it served as a reminder to the coder, thereby reducing the amount of rewinding as well. Hence, although the coder had to stop more frequently, the tool actually reduced the time consuming action of rewinding the video to repeatedly listen to unsure sentences in normal sessions.

Figure B.1: The semi-auto transcribing software
Using this semi-auto transcribing tool developed, all interview videos were transcribed into a table, with three columns: timestamp, transcript and speaker. Since precise timestamps of exact sentences are not significant in this data, shorthand of [...] was used, when the coder did not input the start of a particular line, and it was assumed to be a time in continuation of the last line. This technique significantly helped to reduced time for coding caused by strict time-coding sentences as suggested in some literature (e.g. minute by minute breakdowns).
Appendix C

Case Study: Attempts at Parametric Reduction

C.1 Introduction

After the first two studies, we have a better understanding of differences between designers in terms of design processes and use of computational design tools. We decided to investigate more in details for the design exploration behaviors in a laboratory setting. In particular, we were interested in the choice overloading phenomenon, which can happen at various stages of design as reported by design professionals in Study 2, regardless of whether computational tools are used. However, it’s noted choice overloading effects can be exaggerated when generated computational tools are used simply due to the overwhelming number of generated solutions. We decided to observe a specific design task process in a lab setting with overloaded number of choices.

C.2 The Parametric Model: Residential Apartment

In this case study, we designed a parametric model with 13 parametric entities, including 27 independent parameters to serve the hypothetical design task for a residential apartment (Figure C.1), using Generative Components (Figure C.2). The building design includes three basic components: residential units (grey), vertical circulation spaces (red), and horizontal terraces (green). Each of these components is a parametrically defined rectangular prism. The distance (spread) between the same type of components is defined by a non-linear function: the distance between the same type of components can vary. The points define the spread of units in the structure and are part of a law-curve controller. The size of each component can be parametrically adjusted, and their locations are decided based on the building size constraints. The parameter value ranges used in this case study are shown in Table C.1.
C.3 Selecting a Set of 1000 Alternatives

In this study, the design task is assumed to be in context of the initial conceptual design phase. In this phase, the designer attempts to experiment with different algorithmic options to find the best set of 1000 alternatives as a starting point for further exploration. The presumed goal for the designer is to explore the initial parametric space to search for inspiring forms in concepts. It is likely the designer will continue to refine the constraints and values ranges of the parametric model in later explorations. The number 1000 is chosen as an arbitrarily good number to start - it could have been much more or less depending on personal preferences, however, we are just demonstrating our approach in this case study and chose 1000 as our benchmark.

The purpose of this case study is to experiment with different algorithmic options in a real design task to evaluate whether the proposed approach discussed in Chapter 5 is appropriate. Different dimensions attempted in algorithms are illustrated below.

Alternatives Generation - Structured vs. Random Sampling

During the first step of the task, the designer must generate a limited number of alternatives to start. We tried both structured and random generation in this case. In the structured generation option, alternatives are generated incrementally with a fixed parameter range chosen by the designer, with fixed values or fixed step-sizes. For this case, in order to keep the numbers reasonable for further steps, the designer has to make a deliberate decision to fix some of the less important parameters, varying only 12 parameters of her personal preference. Only 2 possible values are chosen for
the selected 12 parameter, hence results in structure generation of $2^{12}$ i.e. 4096 alternatives via the computer (Sample generations of structured method shown in C.4). In the random generation option, we ask the computer to randomly choose a value in the entire range for all parameters on each new alternative generation (Sample generations of random method shown in C.3). The total number of random generations varies during our experimentation with different algorithmic options for filtering scope, as discussed in later sections. A comparison between 4096 alternatives generated by structure and random options unsurprisingly shows: 1) random generation displays more visual variability in the architectural design, and 2) structure generation is able to reveal the relationships between the visual output and the parameters to someone who has no knowledge about the parameter model itself.

**Bray-Curtis Dissimilarity - Normalized vs. Raw Parameter Values**

We choose to use Bray-Curtis for this case study as part of the exploration. The Bray-Curtis dissimilarity is a statistic commonly used in ecology to quantify compositional differences between two sampled sites based on count data from every species presented at the sites. We are adopting the measure to quantify compositional differences between two generated design alternatives based on parameter values from each alternative. We calculate Bray-Curtis dissimilarity index using raw parameter values (Equation C.1) and normalized (Equation C.2).

The advantage of using the normalized values (scaled normalizations from 0 to 1) in the case study was apparent, as it is more reflective of the visual difference between two alternatives on average. Using only the raw parameter values to calculate dissimilarity resulted in some parameters taking more weights in the calculation due to its large raw value. For example, without normalization, a difference of 2.5 in parameter [length of grey cube] weighs 5 times than a difference of 0.5 in
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Range (Min to Max)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Building Length (Max)</td>
<td>2 - 14</td>
</tr>
<tr>
<td>2 Building Width (Max)</td>
<td>2 - 14</td>
</tr>
<tr>
<td>3 Building Height (Max)</td>
<td>2 - 14</td>
</tr>
<tr>
<td>4 Cube-Grid (X, Y, Z)</td>
<td>2 - 10 for all dimensions</td>
</tr>
<tr>
<td>5 Resident (Gray) (H, W, L)</td>
<td>0.2 - 5 for all dimensions</td>
</tr>
<tr>
<td>6 Circulation (Red) (H, W, L)</td>
<td>0.2 - 5 for all dimensions</td>
</tr>
<tr>
<td>7 Terraces (Green) (H, W, L)</td>
<td>0.2 - 5 for all dimensions</td>
</tr>
<tr>
<td>8 Point01.X, .Y (Controller)</td>
<td>0 - 1 for all dimensions</td>
</tr>
<tr>
<td>9 Point02.X, .Y (Controller)</td>
<td>0 - 1 for all dimensions</td>
</tr>
<tr>
<td>10 Point03.X, .Y (Controller)</td>
<td>0 - 1 for all dimensions</td>
</tr>
<tr>
<td>11 Point04.X, .Y (Controller)</td>
<td>0 - 1 for all dimensions</td>
</tr>
<tr>
<td>12 Point05.X, .Y (Controller)</td>
<td>0 - 1 for all dimensions</td>
</tr>
<tr>
<td>13 Point06.X, .Y (Controller)</td>
<td>0 - 1 for all dimensions</td>
</tr>
</tbody>
</table>

Table C.1: List of all parametric entities and values used in the Generative Components model. Value ranges for each parameter is listed on the right.

However, [x of point 1 controller]. But after normalization, they will be the same 0.5, which is a more accurate reflection of the visual difference on the parameter differences.

\[
D_{ij} = \frac{\sum_{n=1}^{N} |P_i(n) - P_j(n)|}{\sum_{n=1}^{N} P_i(n) + \sum_{n=1}^{N} P_j(n)}
\]  
(C.1)

where:

- \(D_{ij}\) is the Bray-Curtis dissimilarity between alt. \(i\) and alt. \(j\)
- \(N\) is the total number of parameters shared between alt. \(i\) and alt. \(j\)
- \(P(n)\) is the \(n\)th parameter value in the alternative

\[
ND_{ij} = \frac{\sum_{n=1}^{N} \frac{|P_i(n) - P_j(n)|}{P_{max}(n) - P_{min}(n)}}{2N}
\]  
(C.2)

where:

- \(ND_{ij}\) is the normalized Bray-Curtis dissimilarity between alt. \(i\) and alt. \(j\)
- \(N\) is the total number of parameters shared between alt. \(i\) and alt. \(j\)
Figure C.3: Sample generations (random).

\[ P(n) \] is the \( n \)th parameter value in the alternative
\[ P_{\text{max}}(n), P_{\text{min}}(n) \] are the max and min of the \( n \)th parameter (Table C.1)

Other variations of Bray-Curtis, such as weighted parameters (Equation C.3) can also be used. In addition, extensions to second-order parameters, i.e. parameters derived from initial model parameters (first-order) can also yield better results. For example, total volume of grey cubes is a second-order parameter that can be computed from the 3D model generated based on the 12 parameters listed in Figure 4. Another less obvious second-order parameter is the 2D image itself. Computer vision algorithms can be applied in creative ways to achieve similarity calculations. A simple example is a brute force pixel-based image comparison procedure that compares two images pixel by pixel and gives a percentage difference of the two images as an approximation of the visual difference between the two images. More sophisticated computer vision algorithms can perform feature-based similarity comparisons. Further extensions can be built on second-order calculations for more complex evaluations, such as similarity in sunlight studies.

\[
WD_{ij} = \frac{\sum_{n=1}^{N} \left| \frac{P_i(n) - P_j(n)}{P_{\text{max}}(n) - P_{\text{min}}(n)} \right| \ast W(n)}{2N}
\]  

(C.3)

where:

\( WD_{ij} \) is the weighted Bray-Curtis dissimilarity between alt. \( i \) and alt. \( j \)
$N$ is the total number of parameters shared between alt. $i$ and alt. $j$

$P(n)$ is the $n$th parameter value in the alternatives

$W(n)$ is a user defined weight on the $n$th parameter in the alternatives

$P_{\text{max}}(n), P_{\text{min}}(n)$ are the max and min of the $n$th parameter (Table C.1)

**Filtering Scope - Global vs. Within-set**

Global filtering is performed on fixed number generations with the entire dissimilarity matrix calculated, whereas within-set filtering is performed as each alternative is generated with the already selected set. In this case, a global 4096 x 4096 matrix with each cell containing a Bray-Curtis calculation for the pair, then an average for each alternative, “dissimilarity” against all other alternatives was calculated, a global filtering was done to select the top 1000 most dissimilar alternatives. In the within-set filtering case, dissimilarities are evaluated on a rolling-basis only against those already selected as opposed against all generated solutions. This is beneficial in terms computation time. In this study, we either structurally or randomly select a seed alternative from whatever generation option described in (a), then continue on evaluating the upcoming alternative until 1000 are selected. We only select the upcoming alternative if it has a dissimilarity index above the threshold 0.3 with everything that is already in the selected set. If the upcoming alternative is similar with anything in the set then we discard it and go to the next alternative.

**C.4 Assistive Interactive Visualization**

After experimenting with different methods of generating and selecting the 1000 alternatives, we needed an effective way to evaluate which set of 1000 alternatives is the most suitable set to be used for Study 3 (Chapter 5). An matrix based interactive (Figure C.6) inspired by a very large
zoomed-out Excel table (Figure C.5) is built to assist designers to seek the appropriate set of alternatives. Using this visualization, designers can test different pair-wise dissimilarity or distance metrics (e.g. Bray-Curtis, Euclidean) by: a) displaying an overview visual pattern of the set’s dissimilarity, and b) allowing designers to verify whether the metric values match their expectations of dissimilarity. This approach relies more on designer’s interaction with the alternatives directly rather than an artificial intelligent involved in selection. We believe that this is needed to engage designers who sentimentally value their own judgments over pure machine outputs. Further, based on our observations, we can speculate that the shift balance towards manual selection alternatives using the mixed-initiative approach could be more insightful since the rich dimensions of designers’ perception of similarity, the ongoing changes in designers’ criteria for evaluation, and agility in changing design goals cannot be captured fully by machine learning.

Figure C.5: Inspirations from simply zooming out from Excel Spreadsheet with conditional colouring. Image showing the a red highlighting conditional on if the cell value is greater than 0.37.

Figure C.6: Interface of the dissimilarity matrix. Each cell on the grid is representing the dissimilarity metric calculated. The colors can be set by users as explained in Figure C.7 and C.8 to encode different values of dissimilarity. When the user clicks on a cell (the highlighted and zoomed-out cell here is selected), the two images of the correlated alternatives is displayed on top of the matrix, while showing the alternative code and its dissimilarity value on the bottom of the images.
Figure C.7: Two dissimilarity matrices with the same 4096 alternatives. Left: Three-point gradient scale, where red = very different, white = neutral, green = very similar. Right: Two-point gradient scale, where orange = very different, white = very similar.

Figure C.8: Selected sample pairs to illustrate dissimilarity values based on Bray-Curtis method and their corresponding color representation on the two scales on the right.

C.5 Conclusion

In this case study, we have experimented with different methods of generating and selecting a set of alternatives, and developed a matrix-based interactive visualization to assist the designers to sensually (as opposed to logically) evaluate which set best represent the designer’s needs. This mixed-initiative approach has demonstrated its merits in this case study over a pure machine AI or a pure human approach.
C.5.1 Possible Applications for Computational Design Tools

The experimentations done in this study can be applied into future computational design tools by extending the following aspects:

- more similarity and distance metrics
- more sampling methods and parametric reduction algorithms
- represented with more interactive visualizations

The direction of providing a rich repertoire for both V(isual) and A(analytics) will enable the designers to perform the following tasks without requiring any maths knowledge behind-the-scenes:

- choose a metric that matches their expectations of "dissimilarity" visually
- select a few metrics and "marrying" them together using genetic algorithms (presented as a black-box to users)
- learn about the relationships between the math formulas behind-the-scenes and the visual images to a point where they can write their own formulas eventually

C.5.2 Limitations and Contributions

This case study was a by-product leading into Study 3 (Chapter 5). The scope of the explorations were limited by tight time constraints. Some may question the choice of metric and visualization. We recognize this limitation and are not attempting to generalize the effects of the exact metric and visualization used in assisting designers to generate and select alternatives from a large set. Rather, we hope to present a way to accomplish such task with a repertoire of metrics and visualizations. Further quantitative studies must be done with more extensions to the system (as discussed in above section) to demonstrate the robustness of this approach.

Despite being limited in the above mentioned ways, this study has accomplished its goal to provide the adequate set of alternatives for the next study (Study 3, in Chapter 5), to conduct observations in how designer behave during design exploration when overloaded with computer generated alternatives. We also believe that explorations of this mixed-initiative interactive visualization approach have the potential to encourage designers to capitalize the benefits from generative computational design tools, and thereby produce more divergent designs in a more efficient manner in terms of time and design resources.
Appendix D

Study 3: Pre & Post-Task Questionnaire

The following questionnaires (Pre-task, 3 pages; Post-task, 7 pages) are conducted in Study 3. Please see Chapter 5 for details.
Pre-Experiment Questionnaire

Name/ID: ____________________  Date: __________

1. Please list the design fields you have experience with and rank your expertise in these fields:

<table>
<thead>
<tr>
<th>Design Field</th>
<th>Novice</th>
<th>Expert</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Fine Arts</td>
<td>1 2 3 4 5</td>
<td>[_____]</td>
<td></td>
</tr>
<tr>
<td>2. Graphic Design</td>
<td>1 2 3 4 5</td>
<td>[_____]</td>
<td></td>
</tr>
<tr>
<td>3. Architectural Design</td>
<td>1 2 3 4 5</td>
<td>[_____]</td>
<td></td>
</tr>
<tr>
<td>4. Interface Design</td>
<td>1 2 3 4 5</td>
<td>[_____]</td>
<td></td>
</tr>
<tr>
<td>5. Programming (System Design)</td>
<td>1 2 3 4 5</td>
<td>[_____]</td>
<td></td>
</tr>
<tr>
<td>6. Interaction Design</td>
<td>1 2 3 4 5</td>
<td>[_____]</td>
<td></td>
</tr>
<tr>
<td>7. Industrial Design</td>
<td>1 2 3 4 5</td>
<td>[_____]</td>
<td></td>
</tr>
<tr>
<td>8. Other (________________________)</td>
<td>1 2 3 4 5</td>
<td>[_____]</td>
<td></td>
</tr>
</tbody>
</table>

2. Please mark your expertise level with the CAD software listed below.

<table>
<thead>
<tr>
<th>CAD Software</th>
<th>Never used</th>
<th>Novice</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>SolidWorks</td>
<td>☐</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>Rhino</td>
<td>☐</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>Rhino+Grasshopper</td>
<td>☐</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>Generative Components</td>
<td>☐</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>Maya</td>
<td>☐</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
<tr>
<td>Other (_______________________)</td>
<td>☐</td>
<td>1 2 3 4 5</td>
<td></td>
</tr>
</tbody>
</table>
3. Have you worked on a project that involved designing alternative solutions? If so briefly describe the project, software systems you used, and if you can recall how many alternatives you worked with.

4. Have you used parametric and generative design methods before?
   □ Yes (explain)________________________________________________________
   □ No

5. Have you generated design alternatives by changing parametric value; and if so what were the mechanisms you used to change the values?
   □ Yes (explain)________________________________________________________
   □ No

6. How many numbers of alternatives do you normally find yourself handling comfortably at one given time for a given project?
7. Among the given list of devices below, which ones you use more frequently during design.

<table>
<thead>
<tr>
<th>Device</th>
<th>1</th>
<th></th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Desktop computer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laptop computer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tablet (e.g. iPad, Galaxy etc)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Post-it notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pen, pencil, paper, eraser</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Print outs from computers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rapid prototypes created using computer models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vertical boards (e.g. blackboards)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Others]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Post-Experiment questionnaire

Name/ID: ____________________ Date: _________

1. If this project was assigned to you in real life, how would generating a large number of alternatives effect your conceptual-design phase? Explain how many alternatives would you generate and why?

2. How many design variations would you generate at other phases of design development?
3. Recall your process of exploration during the experiment. Can you explain with your words what decisions you made and how? For example, what did you think when you received all the alternatives in the beginning of the experiment; and how did you derive to the final choices?

4. During the experiment, can you recall how many numbers of alternatives you handled at one time simultaneously? Was this number consistent throughout? Why?
5. How many alternatives you have viewed? How many do you think you have viewed for more than 5 seconds?

6. During the experiment, did you find yourself in a difficult situation when narrowing down possible solutions?

7. At any given time did you find yourself using a specific area of desk or white board in a very specific or meaningful way?
8. What did you mainly used the horizontal surface for?

9. What did you mainly used the vertical surface for?

10. Do you think the purposes for horizontal and vertical surfaces can be switched? If so, how?
11. Do you recall how using horizontal surface affected your design decisions?

12. Do you recall how using vertical surface affected your design decisions?

13. Was there any difference in number of design alternatives you could handle on horizontal versus vertical surfaces? What could be that number for each surface in the context of this experiment?

14. Which surface helped you more in managing your selection process and in what ways?
15. Did you find yourself marking or dividing the same surface into sub-surfaces to accommodate various selection tasks? If so, how?

16. Did you find using the given devices (markers, sticky notes, pen/pencils, magnets etc.) helpful? How?

17. Do you think having this set of devices affected your design decision in any way? How different devices are assisting the process differently?

18. Did you feel you had access to all the devices you wanted to use for this experiment?
19. Imagine that you are presenting your rationale of selection to a group of designers. How would you explain your process of exploration and selection? Please include sketches in your answer.