Re-representation: Affordances of Shared Models in Team-Based Design

Jonathan Edelman and Rebecca Currano

Abstract The use of media within the process of designing new products has not been directed by rigorous research findings. In this chapter a media-model framework is discussed, which categorizes media according to levels of resolution and abstraction. This framework can be used to assess characteristics of various models and as a general guide for discerning differences between media types. Designers can utilize the media-model framework to make informed judgments about appropriate prototyping and modeling approaches within various stages of the design process. New research in the application of media-models to Business Process Modeling (BPM), which traditionally employs electronic media (in the form of complex computer-generated flow-charts) aids in the generation of Business Process Models. This research has resulted in the development of an innovative modeling tool, called Tangible Business Process Modeling, or TBPM.

Keywords Media models · Media cascades · Prototyping · Tangible media · Business process modeling

1 Introduction

Our research investigates the unexamined assumptions that underlie the use of media in current design practice. The object of this type of investigation is to ether debunk best practices or to find a scientific basis for them. Can we find rigorous frameworks in order to make informed choices in the course of product or service development?

The hypothesis that we have examined this year concerns the use of rough prototypes in early product development cycles and has two parts (Frederick 2007; Buxton 2007):

J. Edelman (✉) and R. Currano
Center for Design Research, Stanford University, Building 560, 424 Panama Mall, Stanford, CA 94305, USA
e-mail: edelman2@cdr.stanford.edu; bcurrano@stanford.edu

Principal Investigator: Larry Leifer

H. Plattner et al. (eds.), *Design Thinking: Understand – Improve – Apply*, Understanding Innovation, DOI 10.1007/978-3-642-13757-0_4,

Rough sketches and prototypes yield **paradigmatic** changes in a model and high-resolution renderings and models yield **parametric** changes in a model.

If we can establish these postulates as true, the next step is to develop an instructional framework which can inform intelligent design and implementation of prototyping strategies to improve product and service development.

Our research led us to explore several domains, including cognitive science, design theory and methodology, science and technologies studies, economics, and information technology. We have been fortunate to have the opportunity to work closely over the past year with business process modeling researchers at the Hasso-Plattner-Institute in Potsdam. This partnership enabled us to collaboratively design a tool to assist designers and business managers in applying the framework we developed, and to test it in both Stanford and Potsdam to see if our assumptions were correct. This collaboration was particularly helpful in assisting the transfer of mechanical engineering design practices into the domain of business process modeling. This in turn, deepened our understanding of how shared models work and led us to make adjustments to improve our framework.

2 Media Models and Media Cascades

During the initial phase of our investigation, we looked at the properties of the media that design engineers use during product development. Specifically, we examined the resolution of shared models and the kinds of conversations practitioners reported during development meetings. Observations in the field led us to posit that, while ***resolution*** was a critical factor in unpacking shared models, another factor was at work, which we identified as ***abstraction***.

2.1 Resolution

By ***resolution*** we mean the level of ***refinement*** or ***granularity*** that can be observed in the fit and finish of a shared representation.

Figure 1 shows two shared representations used in the development of a test car at a major university in the United States. The sketch of the car on the left exhibits lower resolution than the CAD model on the right.

2.2 Abstraction

By ***abstraction*** we mean ***amplification through simplification***, or pulling specific characteristics out of context. This includes the notion of deliberately translating

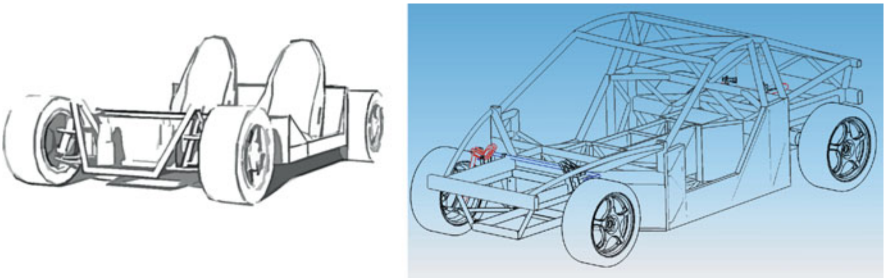


Fig. 1 Low- and high-resolution depictions of an experimental vehicle

Fig. 2 High and low material abstraction in research vehicle

something that is familiar into something *un*familiar. We have observed four classes of abstraction:

1. Material, e.g., material construction
2. Formal, i.e., shape or appearance
3. Functional, e.g., "works-like"
4. Mathematical, e.g., dimensions, optimization

Figure 2 presents an example of two levels of abstraction. The wooden car on the left is more abstract than the steel car on the right. The explicit rationale that went into choosing wood as a material to prototype the car was so designers wouldn't fall into the trap of thinking about how cars are typically designed. The team supervisor, a well seasoned design engineer, felt that using steel would limit choices. This is an example of how abstraction can make the familiar *un*familiar.

2.3 Media Cascades

Hundreds if not thousands of representations are enlisted in the development of new products. We have coined the term ***media-model*** to refer to a ***single representation***

Fig. 3 A media cascade from a student project

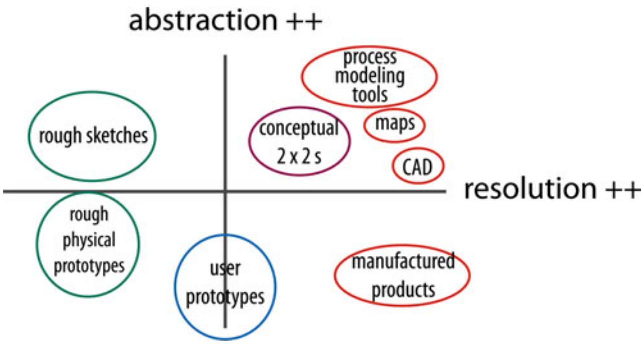


Fig. 4 Media-models framework

in the arc of new product development. Media-models are characterized by the dimensions of resolution and abstraction. We have appropriated Bruno Latour's notion of "cascades of media" to describe the ***sequence of representations*** though which projects develop and unfold in different media during the course of a development cycle, and which we refer to as a ***media-cascade***.

Figure 3 depicts some highlights of a media cascade from a student project, a binding for a snowboard. Shown here are product briefs, rough sketches, rough prototypes, CAD models, functional prototypes, and an actual working model. Rather than seeing these representations as examples of different classes, we see them as different examples of the same class, media models, and we examine how they differ in respect to abstraction and resolution.

2.4 The Media-Models Framework

Figure 4 shows the framework for media models as a conceptual 2×2 matrix. CAD models are both highly abstract and highly resolved. In CAD rendering, specific and actual physical things are reduced to geometric boundaries, or lines, which have no specific material existence. CAD models refer to an entire class of objects, not one real object. In this respect they are highly abstract. CAD models are highly resolved in that they clearly define features and tolerances. There is little or no ambiguity in a CAD model. Instead, design engineers enlist CAD to reduce uncertainty.

Rough sketches and prototypes exhibit low resolution and varying levels of abstraction, depending on the context in which they are used. For example, we consider a sketch to be more abstract than a physical model in the context of designing a physical object. The rationale here is that the three dimensions of the physical object are reduced to two dimensions in the sketch. In the case of the wooden car prototype, the material itself is leveraged as an abstraction to pull out specific design constraints that are invoked by steel.

Note that we consider manufactured products to be highly resolved and not at all abstract. We assume here that the product development has undergone numeric optimization before manufacture. We say that manufactured products are not abstract because they are the actual things.

2.4.1 Completion

Media-models only present a slice of an actual or finished project, and therefore present a ***profile of incompleteness***. A media-model's profile of incompleteness allows design engineers to fill in the presented gaps. Thus, media-models encourage different levels of ***completion*** in order to frame discussion.

Media-models may be classified into three categories - **ambiguous media, mathematized media**, and **hybrid media**. Each class encourages a different kind of completion.

2.4.2 Ambiguous Media

Ambiguous media, such as rough sketches and rough physical prototypes, serve as a scaffold for engineers to fill in the gaps, and are completed as engineers posit many possible formulations of the problem. They are pluri-potential objects, and may express as variants depending on the experience and knowledge of each design engineer who works with them. They encourage ***divergent conversations*** (Fig. 5).

The objects say: I am not the real thing. I am an ephemeral notion.

Fig. 5 Ambiguous media prototype for a communication device

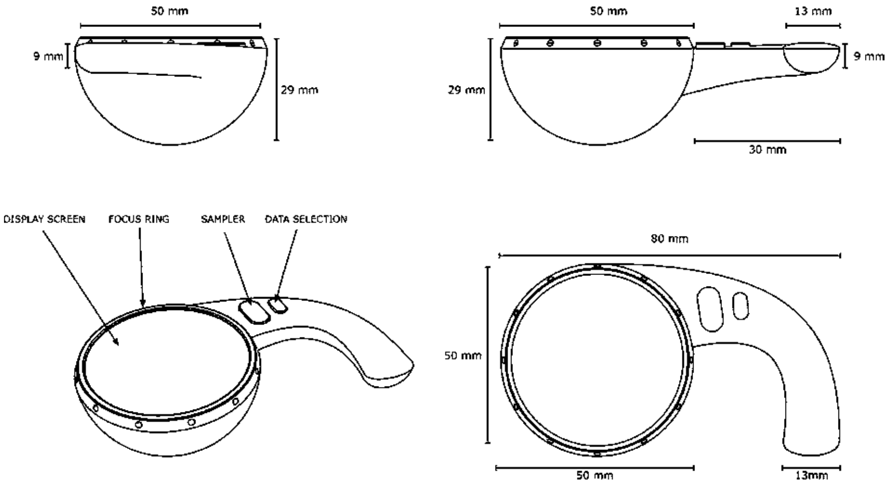


Fig. 6 Mathematized media, CAD model of a device for analyzing material

2.4.3 Mathematized Media

Mathematized media, maps, and highly realistic images are completed through refinement of what is presented. Thus they encourage **convergent conversations**. These media-models present themselves as sacrosanct, and seem to resist substantial changes (Fig. 6).

The objects say: I am the real thing. I am the underlying, unchanging truth of the thing.

2.4.4 Hybrid Media

Hybrid media *allow several kinds of operations and discussions*. Media-models in this category are in the sweet spot for design engineers. They often involve using physical interfaces in conjunction with high-level frameworks. This has proven to allow a flexible exploration of how different elements relate to one another. Design engineers are able to move elements to see how they fit into frameworks, as well as change frameworks to see how they describe phenomena. Hybrid models often involve combinations of different media, such as photographs, drawings, and text. The type of media enlisted in hybrid media has an effect of how the model is completed. Mathematized elements tend not to get changed, while ambiguous elements invite change of the element (Fig. 7).

The objects say: I am about provisional relationships among things.

To summarize:

Ambiguous media-models afford **paradigmatic** shifts.
Mathematized media-models afford **parametric** adjustment.
Hybrid media-models afford **understanding** and changes in **relationships**.

Fig. 7 Hybrid media, physical objects with text narrative on a 2 × 2

We have observed that successful design projects employ many kinds of media-models, or have a *broad bandwidth* of media-models. This is because the varying kinds of media-models engender different kinds of thinking and different kinds of exploration. It is through the translation form one kind of media-model to another that insight is gained, and the project is moved forward.

On the other hand, some media-cascades are characterized by a *limited bandwidth* of media-models. Design engineers may make *more judgments* with six CAD models than with one, but they are making the ***same kind of judgment***, and are engaged in the same kind of thinking. This can also be said of media-cascades constituted of a *single, homogeneous material*, such as white foam core.

3 Cognitive Strategies

It is through the agency of media-models, which serve as *cognitive prostheses*, that various kinds of thinking occur. Insight is gained by the *translation of concepts* into media that embody different levels of resolution and abstraction. Successful product development is dependent on the ability of a design team to employ different *cognitive strategies*.

The underlying assumption here is that by moving the choice of representation around the media-models framework, design engineers can benefit from different kinds of thinking. As mentioned, highly resolved, abstract media is associated with parametric adjustment. Media that exhibits low levels of resolution and high levels of abstraction is associated with paradigmatic shifts. In order to understand these phenomena, we turn to contemporary findings in cognitive science and to an experiment of our own.

Andy Clark has pointed to research that indicates that certain kinds of thinking cannot occur unless subjects' hands actually move (Clark 2008). Clark asserts that much of what we consider to be thinking happens in the hands as well as the mind. Clark's research suggests that thinking doesn't happen only in our heads but that "certain forms of human cognizing include inextricable tangles of feedback, feed-forward and feed-around loops: loops that promiscuously criss-cross the boundaries of brain, body and world" (cf. Clark 2008, p. 129f). In other words, the media itself has an effect on how and what design engineers can think.

Cognitive scientist Barbara Tversky has observed that when presented with rough sketches, experimental subjects engaged in what Tversky calls sketchy thinking (Tversky et al. 2003, 2006), or the ability to think conditionally, or roughly. Other work in cognitive science has investigated the fitness of representations. According to the Cognitive Fit theory, the way the problem is re-presented determines the thinking model applied (Agarwal et al. 1996; Vessey and Galletta 1991). All of this research supports the notion that the kind of media and the characteristics of the media with which people engage have a profound effect on how they think and consequently on the nature of their conversations.

4 Experimental Data

We performed an experiment in order to test some of the assumptions that underlie the media-models framework. We videotaped four teams of three members in two redesign tasks using two different stimuli. One stimulus was a CAD model of a device intended to analyze the properties of material (Fig. 8). The other stimulus was a rough physical prototype of a device intended to project a voice to a specific user (Fig. 9). Instructions were deliberately left vague, in order to see what effects the stimuli would have.

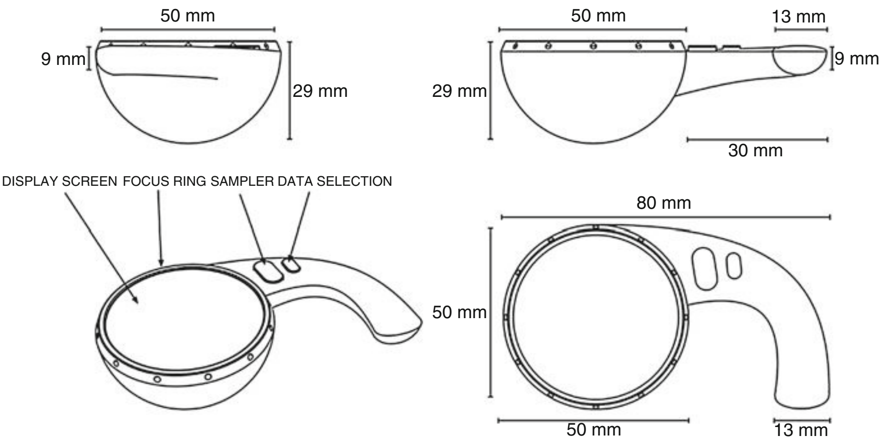


Fig. 8 Experimental stimulus, CAD model

Fig. 9 Experimental stimulus, rough physical prototype

Our instructions for the CAD model were, "Do whatever is necessary to take the model forward. This product enables you to analyze *and identify the material composition of objects*."

Our instructions for the rough prototype were, "Do whatever is necessary to take the model forward. This product enables you to project your voice to a specific target."

4.1 Results

The results were for the most part what we expected. The CAD model generally led to convergent conversations, meaning only parametric changes in the model, or the addition of features. Discussions pertaining to the rough prototype generally led to divergent conversations, which suggested big changes in the model, including adding functionality, as well as frequent additions of features. However, when teams deviated from the norm, we got a close look at how teams innovate and how media supports innovation.

When teams suggested paradigmatic changes in the presence of the CAD model they made lots of rough sketches and prototypes. In a very real way, they covered the CAD rendering with rough sketches. This had the effect of giving the CAD model a cognitive vote, but not a veto. This reinforced our assumptions about the effect of objects on conversations.

4.2 New Insights

We also noted several behaviors that were unique to these teams:

1. Statement of **intention**
2. Asking **process questions**
3. Envisioning **user scenarios**

4. **Enacting** user scenarios
5. Combining metaphors (**It's like X + Y**)
6. Experiencing eureka moments **("Ahhh!")**

4.2.1 Statement of Intention

When teams were able to deviate from the norm of convergent conversations while engaging with a CAD model, they explicitly agreed that they intended to change the model completely. Other teams either never agreed on what they intended to do, or agreed to improve the existing idea. An example of strongly stated intention is, "…[Let's] throw out this design all together."

4.2.2 Asking Process Questions

We noted a significant increase in the number of process-oriented questions in the conversations of the teams who deviated from the norm. Process questions refer to how the team will approach the problem, rather than focusing on the problem itself. An example of this is, "Do we want to make assumptions about whether this is used in the field or in the lab?"

4.2.3 Envisioning User Scenarios

User scenarios differ from *use cases* in that the latter are generic assumptions about a class of users and don't take into account specific circumstances of engagement. An example of a use case would be, "archeologists could use this." The conversations of all teams with both stimuli contained numerous examples of use case. The conversations of teams who deviated from the norm also included numerous depictions of *user scenarios*. User scenarios tend to concern an actual user in a specific situation, often described with rich sensate detail. An example of a user scenario is, "This is so cool that people will want to use it doing anything. They'll use it all the time. They'll be going home and they'll steal it from work…What's in my counter top…"

This insight led us to postulate a mechanism for how radical change to a model occurs. We refer to these changes as "K–C Transits", that is "Knowledge to Concept Transits", invoking Hatchuel and Weil's work on C–K Theory (Hatchuel and Weil 2002) (Fig. 10).

The Cad model may be considered to be an anchor object when it is the single reference point that influences the conversation. When teams generate several user scenarios, they loosen the authority of the anchor object, and the user scenarios themselves become new anchor objects, affording new perspectives that allow the team to make changes in the CAD model, which is now a mutable object.

Fig. 10 Mechanism for K–C Transit

4.2.4 Enacting User Scenarios

Enactment is a phenomenon related to user scenarios, and often occurred in the course of describing a user scenario. Enactment can be observed when team members *act out* the use of an object. This can occur by either pantomiming the action or using a proxy object like a water bottle or a cell phone to represent the object while enacting a scene in which the object is being used. Here again, we found numerous examples of enactment in teams that made paradigmatic shifts with CAD models.

4.2.5 Combining Metaphors (It's Like X + Y)

It's like X + Y involves combining two example metaphors, and seemed to occur in conjunction with enactment. When teams used single instances of metaphor to describe how a stimulus was thought to work, we observed that functional changes would be made to the model. However, when two metaphors were combined, we found that paradigmatic changes in the model occurred. For example, one team combined the metaphor of a scanner with the metaphor of a glove during an enactment and came up with a new notion that was a radical departure from the form of the device in the CAD rendering.

4.2.6 Experiencing Eureka Moments (Ahhh!)

In teams that achieved their explicit intention to change the model, a moment of excitement and recognition occurred. The teams seemed to have co-crafted a new, shared vision. Outbursts of "Oh yeah!" and "Ahhh!" were recorded, and these teams set about hammering out the details of the vision.

5 Tangible Business Process Modeling

The development of **Tangible Business Process Modeling (TBPM)** began in response to information deficits in process elicitation. A year ago, our colleagues at the Hasso-Plattner-Institute approached us in order to collaborate on solving the problem of how to get meaningful information about processes from end users. Software implementations are only as good as the blue prints (process models) upon which they are based, and theirs depend on solid, nuanced end user input. Current elicitation practices rely heavily on end-user interviews, which have yielded less than satisfactory results (see Grosskopf and Veske, Sect. 2): when confronted with a formal business process model and a narrative, which embodied a distillation of the interview, most end users found themselves at a loss. Hundreds of thousands of dollars are typically spent on software implementations of the knowledge contained in process models based on interviews, only to find that important information was missed.

5.1 *The Media of BPM*

When we examined the media (Fig. 11) that was in use during the process model development, we noted that BPM media-cascades are overwhelmingly weighted toward highly abstract and highly resolved media-models (Fig. 12). We were anxious to see if adding media-models with a different profile of abstraction and resolution would help solve the problem. IT was new ground for us, as we had been accustomed to dealing with physical products and services, and not used to translating user input into process models.

In working with business process model researchers, our driving question was, "How can we engender better conversations among the stakeholders (domain experts and process experts) by consciously shifting the BPM media-models around our framework?" We felt that working in a new area was a perfect way to test our framework in a real world arena. In the course of addressing this question, we made unexpected discoveries about media-models and how they worked.

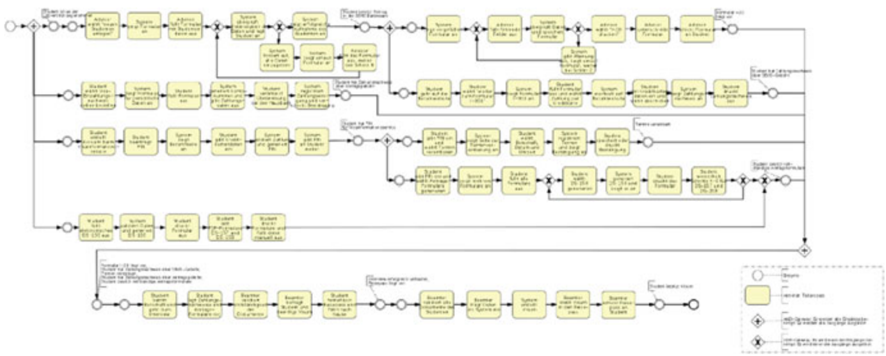


Fig. 11 BPMN process model

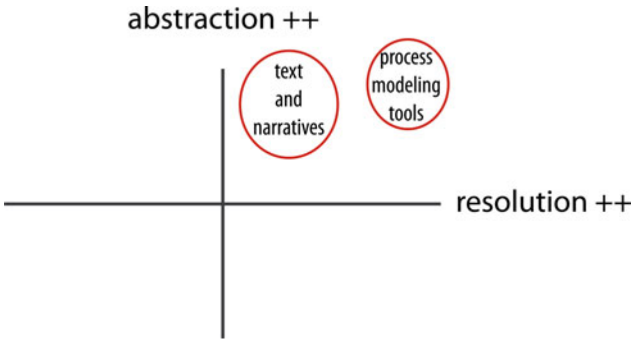


Fig. 12 BPMN media-models

5.2 Intermediary Objects

A second, related question we tackled concerned a close relation to the media-model framework. (Eric Blanco has found that representations serve as intermediary objects Blanco et al. 2007; Boujut and Blanco 2003), which act and are acted upon in the network of design practices and which permit distributed cognition. According to Blanco, shared models may be considered as enlistment devices, either allowing or barring access to collaborative participation. Media that allows collaboration is *open* and media that restricts collaboration is *closed*.

Often, the modeling that is done with software can be considered *closed*, as it keeps control of the model and possible changes in the model in the hands of a few people that are experts with the software tool. Another characteristic of closed models is that they contain little or no explicit affordance inviting change from stakeholders.

The media of BPM has little affordance for direct user involvement. Process experts own and drive the model. As a result, domain experts are left to watch. This means that they have difficulty accessing the kinds of thinking that hands-on work fosters. Thus, our second question became, "How can we create an *open* media in order to give direct involvement to BPM end users?"

5.3 Development of TBPM

We went through several iterations of prototyping strategies in the course of developing Tangible Business Process Modeling, or TBPM (Fig. 13).

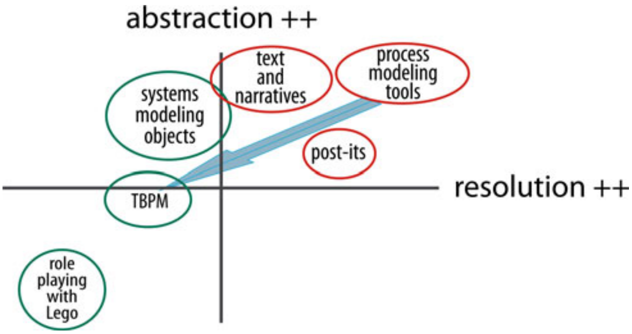


Fig. 13 Media-models used in the development of TBPM

Fig. 14 Role-playing with Legos™

5.3.1 Role-Playing with Legos™

Our early explorations into changing the media of BPM centered around the notion of getting process experts and domain experts to engage in role playing using Lego™ blocks to represent stakeholders and their places of work (Fig. 14). In respect to the media-models framework, this is a move away from high abstraction and high resolution (Fig. 13). While not as concrete as actual enactments, these mediated simulations encouraged players to gain empathy and insight with other players. We also found that the simulation was often cumbersome, encouraging a level of process detail that seemed unnecessary.

Fig. 15 Interview using Post-It® Notes

5.3.2 Post-It® Notes

In another experiment, we enlisted a favorite media of design thinking practitioners, Post-It® Notes (Fig. 15). We found that Post-It® Notes served as an excellent memory aid for domain experts in recalling the steps of their processes. This type of media also provided an object that both the domain expert and the process expert could point to for clarification. One significant shortcoming to Post-It® Notes, we found, was that it failed to frame the elicited process in terms of BPM. This meant that domain experts failed to develop insight into BPM structures, insights which we believe would be central to breaking down the barrier to informed involvement in later stages of process modeling.

Even though Post-It® Notes are easily moveable and rearrangeable, they did not seem to encourage domain experts to express their processes in terms of parallelism or alternatives. Post-It® Notes allowed domain experts to quickly enumerate the steps of their process, but did not lead to greater depth in their understanding. When domain experts were asked if there was anything else they would like to share about their processes as laid out in Post-It® Notes, few if any changes were made.

In respect to the media-models framework, Post-It® Notes are less abstract and less resolved than traditional BPM media (Fig. 13). Our observations of domain experts and their conversations when using Post-It® Notes support our laboratory findings about media and conversations in design teams.

5.3.3 Systems Modeling Objects

In our next iteration, we made a set of acrylic blocks based on Systems Modeling Language (Odum 2004; Meadows 2008). Domain experts and process experts could use dry erase markers and write directly on the acrylic blocks (Fig. 16). Users

Fig. 16 Interview using systems modeling objects

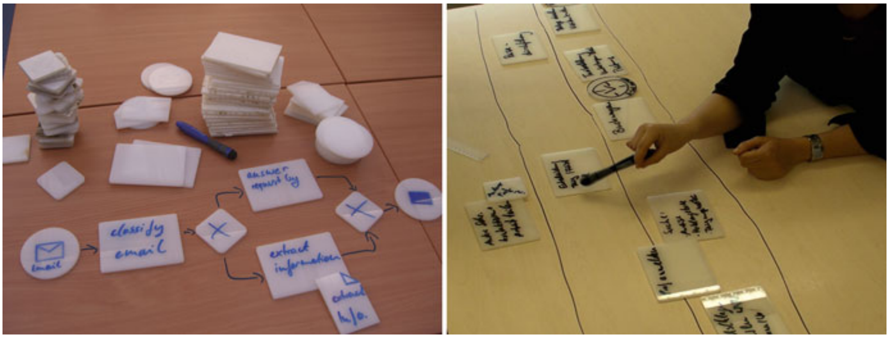


Fig. 17 TBPM elements and interview with TBPM

reported that the pieces were gratifying to handle, and that it was easy to make changes in their renditions of their processes by sliding the pieces around the table. With respect to the media-models framework, tangible systems modeling objects constitute a move towards less resolution and less abstraction than traditional BPM media (Fig. 13).

We found systems modeling rubric to be somewhat arbitrary in respect to BPM. Why not try a tangible set based on BPM Notation? While this may seem obvious in retrospect, it was not a clear choice to begin with.

5.3.4 Tangible Business Process Modeling (TBPM)

In our current instantiation of BPM media, TBPM, we translated four basic BPMN shapes into tangible acrylic pieces (Fig. 17). Virtually all other BPMN shapes could be made form these four basic shapes by writing on them with dry erase marker. With TBPM, the table itself becomes an explicit player, upon which BPM swim lanes are drawn, along with lines connecting activities.

Experimental subjects who had used TBPM elements had the benefits of a memory aid. However, they were able to frame their experience not simply as steps, as with Post-It® Notes, but as a process, which included an understanding of parallelism and alternatives, achieved by placing TBPM elements above one another. The final question, "Is there anything else you would like to add?" led to numerous adjustments and changes, including exceptions to the process they had not yet reported.

When we observed interactions between domain experts and process experts, we found the heightened level of involvement of both parties striking. Domain experts easily grasped fundamental BPM concepts, noting parallelism and alternatives in their processes. At the end of modeling sessions, when asked if there was anything else they would like to share, domain experts dug more deeply into their processes, noting details, exceptions, and making changes.

We were also surprised to learn that interviewers exhibited a higher level of engagement with the domain expert than they did in standard interviews or interviews conducted with Post-It® Notes. We speculate that this was due to two factors: using a shared physical object, and using a domain specific representation.

Interviewers were able to focus on the unfolding of the process without attempting to hold all the disparate pieces in their memory or having to concentrate on writing down responses to questions. Furthermore, because TBPM is domain specific, the burden of interpreting data was lifted, since much of the work of interpretation happened on the table. It seems that our bias toward the end user in user-centered design obscured the fact that there was more than one kind of user in the equation. We realized that the interviewer, or process expert, is a user as well.

In respect to the media-models framework, TBPM is an example of hybrid media, tangible elements on a loosely rendered timeline. While more abstract and more resolved than our first attempt, role-playing with Lego™, TBPM is less resolved and less abstract than traditional BPM media (Fig. 13).

One of the most enjoyable outcomes of our work with HPI was the unexpected learning we gained. The biggest insight showed us the importance of domain-specific instantiations of media-models. Having worked almost entirely in mechanical engineering design, we took for granted that the media-models we used should be situated in that practice. It took some time to realize that we had to create media-models that were specific to BPM. Through framing the *media of TBPM* as a process model, we found that end users could frame their *experience* also as a process with very little training. This made it much easier for them to understand and give meaningful feedback when they were presented with formal process models.

The comparison of TBPM and Post-It® Notes told us that simply having movable objects did not make for meaningful interviews. In order to be effective, shared media needs to be tuned to the domain in which it is situated.

6 Conclusion

While our research is still in the early stages, the media-models framework has been a successful guide to navigating the landscape of shared media. Through the empirical investigation of shared objects, we have brought a modicum of research-based rigor to the perceived best practices of the engineering design community.

The media-models framework has been particularly useful in identifying the "missing media" in BPM practice. As detailed in this paper, traditional BPM relies heavily on highly abstract and highly resolved media, such as flow charts and well-formed narrative. Media of this sort has been shown not to support paradigmatic shifts in the shared model, a necessary component of innovative design thinking. Furthermore, the media of traditional BPM has been "owned" by process experts, which has the effect of barring domain experts from participating in uncovering their implicit processes.

In response to our first question, "How can we engender better conversations among the stakeholders (domain experts and process experts) by consciously moving the BPM media-models around the media-models framework?" we have found that by translating traditional BPM media into media characterized by lower resolution and lower abstraction, we have made a more flexible media that supports exploration and negotiation amongst stakeholders.

As for the second question, "How can we create an *open* media in order to give direct involvement to BPM end users?" TBPM explicitly delivers control of the model into the hands of all participants at the table, allowing the ability to "think with their hands" as they work out what the nature of their implicit process is. TBPM affords moving, rearranging, adding, and putting aside its elements.

7 Future Work

More work remains in the development of a robust TPMN toolkit and methodology. Process improvement is a facet of BPM that is particularly interesting to us. We believe that TBPM will prove to be a strong tool for process improvement among existing stakeholders. However, a broader perspective may serve process improvement. To that end, we are exploring Customer Value Chain Analysis (Donaldson et al. 2006) with the TBPM toolkit. It is our desire to develop media and methodologies which support a *generative* approach to modeling the Customer Value Chain (CVC) in contrast to an *analytic* approach. Though physical embodiment of relations, time, and effort we hope to uncover pain-points for stakeholders. This in turn will allow designers to explore alternatives to the current CVC. The insights gleaned from this approach can be channeled back to TBPM as a framework for making global improvements to process models.

References

Agarwal, R., Sinha, A., Tanniru, M.: Cognitive fit in requirements modeling: A study of object and process methodologies. Journal of Management Information Systems 13(2) (1996) 137–162

Blanco, E., Grebici, K., Rieu, D.: A unified framework to manage information maturity in design process. International Journal of Product Development 4(3) (2007) 255–279

Boujut, J., Blanco, E.: Intermediary objects as a means to foster co-operation in engineering design. Computer Supported Cooperative Work (CSCW) 12(2) (2003) 205–219

Buxton, W.: *Sketching User Experiences: Getting the Design Right and the Right Design*, 2007, Morgan Kaufmann, San Francisco, CA

Clark, A.: *Supersizing the Mind*, 2008, Oxford University Press, Oxford, UK

Donaldson, K.M., Ishii, K., Sheppard, S.: Customer Value Chain Analysis, in Research in Design, 2006, 16: 174–183, London

Edelman, J.A., Leifer, L., Banerjee, B., Jung, M., Sonalkar, N., Lande, M.: Hidden in Plain Sight: Affordances of Shared Models in Team-Based Design, International Conference on Engineering Design, ICED'09, Aug 24–27, Stanford, CA, USA

Frederick, M.: *101 Things I Learned in Architecture School*, 2007 MIT, Cambridge, MA

Hatchuel, A., Weil, B.: "C-K Theory: Notions and Applications of a Unified Design Theory", 2002 Proceedings of the Herbert Simon International Conference on Design Sciences

Odum, H.T.: *Ecological and General Systems: An Introduction to Systems Ecology*, 1994, University Press of Colorado, Niwot, CO

Meadows, D. H.: *Thinking in Systems*, 2008, Chelsea Green Publishing Company, White River Junction, Vermont

Tversky, B.: "What Does Drawing Reveal About Thinking?", 2006 Lecture for *Visualizing Knowledge Seminar*, Stanford University

Tversky, B., Suwa, M., Agrawala M., Heiser J., Stolte C., Hanrahan, P., Phan D., Klingner, J., Daniel, M.-P., Lee, P., Haymaker, J.: Sketches for Design and Design of Sketches, 2003, Stanford University, Stanford

Vessey, I., Galletta, D. Cognitive fit: An empirical study of information acquisition, Information Systems Research 2(1) (1991) 63

The Co-evolution of Theory and Practice in Design Thinking – or – "Mind the Oddness Trap!"

Julia von Thienen, Christine Noweski, Christoph Meinel*, **and Ingo Rauth**

Abstract In Design Thinking, theory and practice are closely interconnected. The theory serves as a blueprint, guiding companies in general and design teams in particular through the design process. Given such a close interrelation of theory and practice, we argue that Design Thinking research needs to be set up in a particular way too. This setup ties in with Design Thinking process models: To attain ever more befitting design solutions, prototypes are supposed to be tested and refined. Correspondingly, Design Thinking research should help to test and refine theory elements of Design Thinking. Researchers may serve as "dialogue facilitators," aiding the community of Design Thinkers to intensify their "dialogue" with empirical reality.

To provide reliable data on issues of central concern, we have tested experimentally two widely held convictions in the field of Design Thinking: (1) Multidisciplinary teams produce more innovate design solutions than monodisciplinary teams. (2) Teams trained in Design Thinking (by the D-School) produce more innovative solutions than untrained teams. In addition, degrees of communication problems were assessed. While both "multidisciplinarity" and "D-School training" have been associated with more unusual design solutions, with respect to utility a different picture emerged. Thus, hotspots have been identified that may stimulate some productive refinements of Design Thinking theory.

1 From Design Thinking to Design Thinking Research

How should teams approach design challenges? What do students need to learn to tackle design challenges successfully? With increasing frequency, *Design Thinking* is called upon to help answer these questions. Used by multiple big companies such as SAP, P&G, IDEO or GE Healthcare, accompanied by a lot of media attention and

J. von Thienen (✉), C. Noweski, C. Meinel, and I. Rauth
Hasso-Plattner-Institute, Campus Griebnitzsee, P.O. Box 900460, 14440 Potsdam, Germany
e-mail: e.valuate@hpi.uni-potsdam.de; meinel@hpi.uni-potsdam.de

* Principal Investigator

H. Plattner et al. (eds.), *Design Thinking: Understand – Improve – Apply*, Understanding Innovation, DOI 10.1007/978-3-642-13757-0_5,

propelled by an increasing number of training institutions, Design Thinking seems on its way to become the state-of-the-art innovation method. And yet, we understand only little about what really matters for it to be successful.

In the armchair we may think about these issues, but many crucial questions will remain unanswered. Those who truly wish to know will have to confront the real world: Careful empirical analyses are in place! With this thought in mind, we decided to make a real job of it – and put fundamental assumptions of Design Thinking to an experimental test.

Naturally, in the booming, buzzing field of Design Thinking there are innumerable aspects that warrant careful scientific investigations. Of course, one might just cherry-pick some questions, selecting the issues according to personal interests. Yet, the research ought to take into account the interests of people working in the field as well, or shouldnt it? So, we made it our first empirical research task to scan in a somewhat broader fashion the interests, hopes and worries of experts in the field. But sure enough, there was some trouble ahead: While the term "Design Thinking" seems to allude to a common set of practices and a common theoretical matrix, the experts held ready an astonishing variety of understandings. What does that imply for the task of testing empirically central assumptions of Design Thinking theory? Our answer will be an outlook on research endeavours particularly designed to match the characteristic relation of theory and practice in Design Thinking. It will be the basis we start from and return to in our experimental work.

2 Experts Revealing What They Think About Design Thinking

In the winter of 2009, we had the opportunity to speak to a number of Design Thinking experts and conducted a series of guideline interviews of about $1\frac{1}{2}$ h each. In this context, we wish to thank once more members of IDEO, the Design Services Team of SAP, design consultants from Procter & Gamble and Palm as well as members of the staff and teachers of the Design Schools in Potsdam and Stanford. The interviews focussed on three major issues:

1. The definition and understanding of Design Thinking (the process and its methods) as well as prototypical conflicts in Design Thinking projects
2. needs regarding the work environment and tools
3. successful team orchestration and its specific needs

Key insights were synthesized using storytelling and clustering techniques within the project team. Papers have, or will be published on each of the topics. Here, we shall briefly review those issues that helped to shape our further approach within the *HPI* research program.

What stroke us as most momentous for the whole enterprise of Design Thinking research was the grand variety of understandings across experts in the field: The interviewees did not convey a common understanding of Design Thinking. They specified differing process models and named differing methods as crucial elements of the design process.

We found, for example, opposite beliefs regarding the question whether design work should be outsourced or not. According to some experts, design teams need to work outside of common business contexts to avoid being "captured" in their routines. These experts argue that creative freedom needs to be maximized. Ideally, the development of new design ideas should therefore be outsourced. Other leading experts prefer integrative approaches where managers set up teams by bringing together employees from different departments. This way, a single team may attend a project from the earliest up to the latest stages. While different departments are responsible for different steps in the design process (e.g., idea generation versus final implementation), representatives of all departments are joined in the responsible design team right from the start.

To mention another point of divergence, some experts highlight the pivotal importance of individual genius. Others believe, however, that individual genius is comparably unimportant when it comes to predicting the success of a design project. Instead, they say, teams need to be assembled according to sophisticated theories so as to combine particularly "matching" characters and competences.

Interestingly, the experts did not only differ in the concrete approaches they preferred. They explained their understanding of Design Thinking on different scales and reflected upon differing academic discourses. Obviously, there is no common set of beliefs (yet) associated with Design Thinking. Rather, there are differing lines of debate as well as differing practices. To what extent we should strive to bring them together is an interesting question by itself.

Apart from considerable differences in the general understanding of Design Thinking, there were – fortunately! – a number of important commonalities too. Without any such visible connecting factors it would be hard to see how Design Thinking could be studied as a collective enterprise.

A strong focus on **user needs** is considered essential across the board and the aim of true **innovation** is a shared concern. Design teams should not just head for quantitative improvements (such as devising a memory stick with yet more storage capacity, applying well known technologies). They should also be able to bring about qualitative improvements (e.g., by devising new technologies that are more potent or by developing solutions that make memory sticks superfluous altogether). That is, design teams should reconsider initial design challenges ("reframing"): They should try to understand what the users' true needs are. Then, they should consider a whole variety of approaches, including (and quite essentially so) uncommon ones, the so called "wild ideas." In a continuous dialogue with the users, a solution shall finally be worked out that suits the users' needs particularly well.

Another aspect that many Design Thinkers view as central is the academic diversity of design teams. Commonly, **multidisciplinarity** is considered a good choice. Teams are supposed to be academically diverse so that they may integrate impulses from many different domains. It is assumed that multidisciplinarity is particularly well-suited to foster true innovation.

Next to multidisciplinarity, other factors are thought of as crucial for team performance too. In particular, many interviewees stressed the importance of a positive **communication culture**.

In sum, the experts named a number of common concerns. But, strikingly, they did not sketch out a common theoretical matrix associated with the term "Design Thinking." This is a finding that should occupy us! Given the cloudy theory structure of Design Thinking, what are we to expect of Design Thinking *research*?

3 Telling Differences, Illuminating Parallels

Traditionally, theories are considered to be systems of axioms: There are a couple of fundamental propositions from which everything about the field of interest may be deduced. When a scientist refers to "the theory," he refers to its set of axioms. Correspondingly, accepting a theory means to accept "the axioms." With this classical picture in mind, there seems to be something quite worrisome about Design Thinking. If it is a theory – or builds on a theory – where are its axioms? As became all too clear in the expert interviews, there is no common set of propositions that Design Thinkers accept in virtue of their expertise. There are *some* shared convictions that may be understood as guiding theoretical ideas. But, they certainly do not cover the whole domain of interest. Apart from that, rather than there being fundamental assumptions, there are shared *centres of concern*: Usability, multidisciplinarity, unusualness ("go for the wild"), reframing of original tasks – to name some in a random order. Experts occasionally disagree as to how important each issue is in differing project phases. But they routinely monitor and discuss them. Now, what does the lack of a classical theory-structure mean for Design Thinking? Is it non-professional after all? Is it in such an early stage of its development that it has not even managed to produce a meagre axiomatic system?

We, in contrast, believe the "axiomatic system" is a misguided ideal for Design Thinking. There are good reasons for the open theory-structures that characterize Design Thinking today. These open structures are sensible, but nonetheless they may – of course – be improved. To see how the structures make sense and what likely aims there may be for improvements, it seems a good idea to scan the academic field for domains with similar challenges.

Musicology, for instance, does have some interesting parallels to Design Thinking. First of all, its subject is something *productive* and *creative*: Musicologists study pieces of music and their composition just like academic Design Thinkers study design solutions and their coming about.

When looking at – say – pop songs, music theory serves a dual function. On the one hand, it *describes* songs. On the other hand, by working out and comparing song patterns the theory provides a *blueprint* how songs may be composed (successfully). For example, there typically is an intro, then strophes and the chorus alternate, there are bridges, breaks and, finally, an ending. Longer instrumental interludes are typically placed in the second half of a song, not the first.

Yet, such a scheme is not enough for a song. Individual musicians have to fill in the blanks. Novices in particular may profit from following strictly the blueprint they are given. But experts (or: visionaries) may produce masterpieces by breaking the rules. Some of the time, they thus establish new patterns that other musicians will use fruitfully in the future.

In Design Thinking, things are not all that different. Design Thinking theory serves a dual function as well. It helps to describe and analyse design projects (e.g., does reframing happen at some point? What does the team do and when to ensure usability?). Design process models convey standards as to which phases there are and in which order they be put. They also encompass methods that may be invoked.

When a design team orchestrates its own project, it may well profit from given schemes. But sure enough there are blanks to fill in. (For instance, "Here we are in the research phase. We have methods *A* through *H* at our hands. Which shall we pick? How exactly shall we proceed?")

As Design Thinkers grow more and more experienced, they may identify circumstances in which unconventional procedures seem more promising than standard ones. Out they move of common schemes. They break the rules! If this happens, it is an interesting case for Design Thinking theory. Such a "breaking of rules" should not be generally damned. It is a precious test case. Maybe it fails. But if it doesn't, Design Thinking theory hits on an alternative whose potential is yet to be explored.

The parallels between musicology and Design Thinking illuminate two important issues that we need to keep in mind to avoid working towards an inadequate theoretical ideal.

The co-evolution of theory and practice. According to the classical understanding, a theory is true if it describes the empirical world correctly. An unbridgeable gap separates theory and world. Changing the theory will not change the world.

In the case of Design Thinking, as in the case of musicology, the gap is being crossed all the time. Since the theory provides blueprints to practitioners, a change in the theory is likely to change the empirical world itself. Theory and practice co-evolve. In consequence, the question of whether or not Design Thinking theory is true does not "function" in a conventional way. In many respects, Design Thinking theory may be true for trivial reasons: Because it serves as a scheme according to which practitioners proceed. Truth is cheep to have for Design Thinking theory in these regards. And truth does not suffice.

Consider the two claims:

(a)	The theory is true.	True or false?
(b)	The theory is (most) useful.	True or false?

Conventionally, scientists ask whether claim (a) is maintainable. In the case of Design Thinking, claim (b) seems to be the more fundamental, the more demanding. It is the one whose correctness calls for rigorous empirical investigations.

Since theory and practice *are meant* to co-evolve, empirical evidence for a lack of utility will not (and should not) lead to the rejection of claim (b). Instead, careful analyses need to follow. Design Thinking theory – in particular: aspects of its process model – may have to be modified to become ever more useful.

The researcher as a dialogue facilitator. What is the second issue we may – and should – learn from the parallels between musicology and Design Thinking? In our understanding, one more point is particularly important for a proper setting of goals. The example of musicology teaches us how fruitful it can be to have both at the same time: An overall-open theory structure that may seem cloudy – yet a rigorous precision in analytic conceptions.

On the one hand, it is clear that there are many ways to produce felicitous pieces of music; and there are different music styles that may be just as appealing. In this sense, it would be detrimental if musicology would specify one single theoretical matrix according to which music ought to be produced. Musicological theory needs to be open; it needs to be able to handle plurality and to incorporate new developments that the future will (hopefully) bring. This openness in theory structure does not, however, imply that it is necessary or helpful to work with cloudy concepts and claims. For example, think of notes and rhythms that do a marvellous job in documenting and structuring something as elusive as played music! (Do you think you could come up with just *two concepts* such that whole design projects could be reconstructed on their basis? If you have some spare time, maybe sitting in a bus or plain, why not give it a try?)

The aim of potent and precise analytic conceptions – despite of an overall open theory structure – is, we think, an excellent target for Design Thinking as well. While it is clear that Design Thinking theory needs to remain open to allow for new developments, we should still strive to refine our analytical conceptions so that they be ever more potent systematizing factors. We should also try to learn more about our individual versus collective claims – and how well they are substantiated.

With this background understanding, we feel that some rather peculiar role befits us, the researchers. We wish to serve as dialogue facilitators: We wish to help Design Thinkers enter in an intense dialogue with empirical reality. What concepts, what assumptions work well, which do not work all that well yet? The research ought to put Design Thinkers in a position to sharpen their vocabulary and their fundamental beliefs in a way that makes them ever-more adapt to reality, ever more fruitful.

4 Preparing a Look Behind the Curtain: Specifying Hypotheses

As there is no written out axiomatic system in Design Thinking that specifies crucial assumptions one after the other, it is the researchers' first job to pin down crucial beliefs in the field. Our take in the last year was this: In general, it is assumed that Design Thinking fosters innovation. After all, Design Thinking is supposed to be an innovation method (or even: *the* state-of-the art innovation method). So, people who have been trained in Design Thinking should produce more innovative solutions than people who have not been thus trained.

Of course, there are multiple institutes who offer Design Thinking education. As the *Design Thinking Research Program* in Potsdam and Stanford enjoys a close cooperation with the D-Schools in Potsdam and Stanford, the Design Thinking education we shall look at will be a D-School training. Our starting hypothesis may thus be formulated more specifically: It is assumed that D-School trained teams produce more innovative solutions than teams without this training. Additionally, to consider one rather confined factor, we shall test the widespread belief that multidisciplinarity enhances innovation. If the belief is correct, multidisciplinary teams produce more innovative solutions than monodisciplinary ones on average.

While the two hypotheses concerning D-School training and multidisciplinarity are viable starting points, they need to be further refined. In particular, "innovation" is such an abstract notion that it is too remote from potential measurement operations. In such a case, it is usually a good idea to break the abstract concept down into disparate factors that may be assessed more easily. This is our take:

A design solution S_1 is considered more innovative than a solution S_2 if S_1 is more unusual as well as more useful than S_2.

Given this clarification of what "innovative" means, both of the starting hypotheses split into two more specific claims. These are the assumptions regarding D-School education:

1. D-School trained teams produce *more unusual* solutions than teams without this training.
2. D-School trained teams produce *more useful* solutions than teams without this training.

Accordingly, two hypotheses may be formulated concerning multidisciplinarity:

3. Multidisciplinary teams produce *more unusual* solutions than monodisciplinary teams.
4. Multidisciplinary teams produce *more useful* solutions than monodisciplinary teams.

While there are ample reasons to believe that multidisciplinary teams will indeed produce more innovative solutions than monodisciplinary ones on average, there is at least one notable reason to believe the opposite – and it may be fruitful to consider these reasons distinctly.

Experts who have been trained in the very same way of analyzing and approaching a subject matter are likely to invoke the strategies they are all used to when working on a new problem. Whatever work strategies are being used, by and large they pave the way for some particular type of result while detracting from other options. For example, imagine a team of chemists and a team of classical philologists who are to analyze a painting. While the chemists might take tiny samples of the paint and find out which material components have been used, the philologists might identify a scene from Greek mythology and reason backwards to the exact literary sources the painter had been exposed to. Given the specialized knowledge and training of the experts, there seems no way that the philologists could hit on the work results that chemists get and vice versa. Limiting oneself to a fixed set of (common) work strategies usually means limiting oneself to particular types of (common) results. In multidisciplinary teams, however, the approaches that team members are familiar with are likely to differ. Thus, there will be no immediate way of setting about the task. Rather, team members will have to (re-)consider the approaches they find convenient. In bargaining how to move on, they will have to detach themselves from common practices – melding, merging, blending the strategies they know in a way that seems appropriate in the context of their current challenge. The broader the domain of strategies experts are willing to consider, the broader is the domain of results that their team may obtain. Insofar as new approaches are tried, the odds increase that something rather unusual results. Thus, it seems likely that multidisciplinary teams produce more unusual results than monodisciplinary teams.

Regarding the second facet of innovation – usefulness – multidisciplinarity may be all the more advantageous. After all, the development of useful solutions depends upon knowledge, e.g., knowledge concerning the situation of users or knowledge about technical options for realizing some particular idea. Imagine experts who are equally well trained. Clearly, if they are all trained in the very same domain, the knowledge their team disposes of is rather limited compared to the knowledge of a team whose members differ in their fields of expertise. Thus, multidisciplinary teams seem better equipped for developing useful solutions.

Yet, at the same time, there is a reason to believe that, on average, multidisciplinary teams will produce less innovative solutions than monodisciplinary ones. Why that? Even if multidisciplinary teams have a greater potential for innovation, communication problems might hinder them. It seems reasonable to expect that communication will be more challenging in multidisciplinary than in monodisciplinary teams. Just as people with differing academic backgrounds have been trained to use different strategies when approaching a problem, they have also been trained to use different concepts. The words they use may differ, the categories by which they sort things in the world may differ and the implications associated with one or the other categorization may differ as well. If design teams are unable to work out a common conceptual ground, they may not be able to make good use of the wide-ranging expertise of their team members. Thus, we decided to consider a fifth hypothesis that may shed some light on important team processes in the design process:

5. Multidisciplinary teams experience more communication problems than monodisciplinary teams.

At the same time, D-School training might well make a difference with respect to communication success. D-School trained team members might – or rather: they should – be able to handle potential communication problems, whether or not working multidisciplinarily. After all, it is assumed that they are particularly apt for design work. Thus, they must not be thwarted or halted by potential communication obstacles. A sixth and final hypothesis is therefore:

6. D-School trained teams experience less communication problems than teams without this training.

5 Why Experiments Matter

As preliminary considerations have been formulated, a choice needs to be made as to how the subject matter shall be tackled empirically. In principle, two alternatives are available. Investigations can be experimental or non-experimental. Both approaches have their advantages as well as their disadvantages. The experimental method has been devised to fade out or "oppress" all the factors potentially relevant to an outcome except for those factors whose influences are to be investigated (as specified by the hypotheses). Thereby, the relationship between the factors that one takes interest in becomes maximally clear. But, naturally, one doesn't find out anything about the other factors (not addressed by the hypotheses) that one is at such pains to fade out in the experimental setting. In non-experimental studies, on the other hand, one may explore all the facets of real-life situations in their full booming buzzing mix-up. Thus, you may come to consider aspects you would never have thought about in your office armchair, extrapolating from the data hypotheses as to how they *might* be interrelated. Yet, whether these putative causal relations truly exist, one cannot really tell.

In our case, factors have been selected that are of primary interest. The crucial question is whether or not they are causally related. If D-School training and multidisciplinary actually do enhance innovation (as is hypothesized), a hook-up question may be how strong their effect is. These are questions to which experiments alone provide thoroughly compelling answers.

6 The Challenge

In every experiment, the setup requires thorough considerations as it sets the upper limit of what can be found out. In our case, a challenge needs to be formulated concerning a topic that all participants are about equally familiar or unfamiliar with. Otherwise, some teams might dispose over a lot of knowledge regarding the subject matter right from the start as some members would be experts, while other teams would have laypersons only. Regardless of whether one believes that teams profit from an expert (due to their knowledge) or whether one considers experts as a threat to innovation (because they might act as rigorous sensors), the teams with versus without experts would not be working under comparable conditions. Let's assume that, in the end, the presented solutions actually differ in their quality. These differences could not be clearly attributed to the factors of multidisciplinarity versus monodisciplinarity or D-School training versus no such training if the teams had differed in other respects as well, such as expert knowledge versus no such knowledge.

In addition, the scope of the challenge should be somewhat grand, or at least not minute. It should be "open" enough so that it would be possible to come up with a technical or a social solution or an artistic or political or yet other type of solution. A related demand is that there should be the possibility of using knowledge from diverse fields. If, on the other hand, only people with one particular academic training could complete the task (e.g., implement a certain computer algorithm), this would probably forestall successful Design Thinking right from the start.

The challenge that was chosen to meet these needs was this: Come up with something that helps traumatized people to manage their everyday lives!

Indeed, the participants of our experiment (40 students) indicated that their pre-experience with the subject matter, trauma, was basically negligible. For example, no one had ever been a practitioner in the field or had had a considerable training in the domain. Only one student had ever encountered the subject matter in her university studies.

7 Operationalization or: Let's Be Concrete!

Now that a challenge has been specified the question of how to asses, how to "measure" the attributes of interest needs to be considered. Each team will present its suggestion for how to help traumatized people. What is to be done so that reliable measures result, i.e. estimates of the unusualness of each solution?

When invoking numbers in every day life, we often ask questions about concrete things. For example, how many eggs are left in the fridge? In cases like these, we may start counting right away. In our study, on the other hand, the factors of interest are quite abstract. This does make a difference for the procedure of assessing or "measuring" those factors. How is one to count the unusualness of a design solution, for instance? Obviously, some further steps need to be taken.

In order to assess abstract factors they need to be *operationalized*. The question to be pondered is this: Given the context of your particular study, what could you observe straightforwardly to find out about the factor(s) of interest? Your task is to find concrete entities that one can look at to arrive at reasonable statements about the abstract notions of interest.

In the setup of an experiment, the operationalization is a crucial step. If one's operationalization is unconvincing, one's data will fail to bear on the issue that one sets out to investigate! Thus, in the case of our experiment as well as in general, we want to invite you to take a very careful look at the operationalizations: What do people (we) actually observe when they (we) make claims about highly abstract matters? Is the step they (we) take from observed entities to theoretical entities actually warranted? In our case, on the level of theory there are five factors of interest: (1) D-School training, (2) academic diversity, (3) the unusualness of design solutions, (4) the usefulness of design solutions and (5) communication problems.

While the factors (3)–(5) truly call for discussion, for reasons of completeness we shall mention the first two as well. There was a very convenient way of assessing the academic background of participants: We basically asked them. In the case of Design Thinking experience we consulted official lists of D-School trainees and alumni.

What is "unusual"? While the "unusualness of a design solution" is too abstract to be looked at and counted directly, we may ask people questions and attain concrete answers, counting how many times particular replies are given. To arrive at a pertinent question, the following consideration seems reasonable: In the context of our experiment, a group presents an unusual solution if the other teams (who have worked on the same challenge, after all) failed to consider that particular possibility when discussing options for helping.

In the course of the experiment, every team has to present its solution. All the participants need to fill out a questionnaire including the following question – regarding each single presentation (of the other groups):

Item 1 Has the presented solution been discussed in your group as well?

- ☐ Yes, exactly in this form (1)
- ☐ Yes, in about that way (2)
- ☐ More or less (3)
- ☐ No, but that may have been a coincidence (4)
- ☐ No, we would never have hit on it (5)

The brackets show our coding. Thus, the statistical values obtained range from 1 to 5. Greater values indicate a greater degree of unusualness.

Of course, the participants of our study are not the only people to ever think about how one could help in the case of traumatisation. There are experts in the field, trauma therapists in particular, whose job it is to help traumatized people. In addition, there are people who have suffered a traumatisation, of course. They too may have thought about options for improving their situation. Accordingly, these experts shall be contacted, introduced to one design solution after the other and asked a question quite similar to *item 1*:

Item 2 Have you ever considered this option for helping before?

- ☐ Yes, exactly in this form (1)
- ☐ Yes, in about that way (2)
- ☐ More or less (3)
- ☐ No, but that may have been a coincidence (4)
- ☐ No, I would never have hit on it (5)

Again, values range from 1 to 5. Greater values indicate a greater degree of unusualness.

What is "useful"? While the design teams may contribute information regarding the unusualness of a design solution, they are hardly in a position to specify utility. Of course, members of design teams can say something about *what they think* how useful their solution is (and we did ask them this question). Yet, whether or not a tool is actually helpful is not decided by the developers but by the users. In our context, the users are traumatized people or therapists who work with traumatized people. (Many teams actually developed tools that would aid the therapists in helping their clients.)

To attain judgements of how useful each solution is experts have been asked the following question:

Item 3 What do you think, how helpful is this approach for the target group?

- ☐ Very helpful (5)
- ☐ Quite helpful (4)
- ☐ Somewhat helpful (3)
- ☐ Barely helpful (2)
- ☐ Not helpful (1)

Again, values range from 1 to 5. Greater values indicate a greater degree of usefulness.

When working with operationalizations, disposing over a second estimate for each factor of interest is commonly quite advantageous. It helps you check whether the numbers you attain actually represent what they are supposed to. If two different indicators of the very same factor point in the same direction this gives you some (further) evidence for their working properly. If, on the other hand, indicators for the same subject matter point in different directions, this is ample evidence for there being something wrong with your assessment procedure(s). Thus, a second item was formulated that ought to cap onto the factor "usefulness."

Item 4 Which approaches should be realized by all means?

Please mark up to five approaches!

Marked (1)
Not marked (0)

Again, the brackets show our coding. Values range from 0 to 1. Greater values indicate a greater degree of usefulness.

How to assess "communication problems"? Communication problems, of course, would have to be estimated by the team members and not by the experts (who were contacted after the experiment). At the end of the workshop, the participants were asked to fill out a questionnaire containing three items to assess potential communication problems.

Item 5 Was it easy or difficult for your group to reach an agreement?

- ☐ Very easy (1)
- ☐ Easy (2)
- ☐ Neither nor (3)
- ☐ Difficult (4)
- ☐ Very difficult (5)

Item 6 Have there been group decisions that you felt uncomfortable with?

- ☐ Not at all (1)
- ☐ Very few (2)
- ☐ Some (3)
- ☐ Several (4)
- ☐ Plenty (5)

Item 7 Have there been communication problems in your team?

- ☐ Not ever (1)
- ☐ Rarely (2)
- ☐ Sometimes (3)
- ☐ Often (4)
- ☐ Very often (5)

Table 1 The constructs of interest and their operationalization

	Variations (Independent variables)		Outcome (Dependent variables)		
Level of theory					
Of interest	Team setup		Innovation		Communication
	D-School training	Academic diversity	Unusualness of solution	Usefulness of solution	Problems
Level of observation (operationalization)					
Who rated			Experts and teams	Experts	Teams
Observable	Statements, list	Statements	Item 1 (team) Item 2 (experts)	Item 3 (aid) Item 4 (choice)	Item 5 (agreement) Item 6 (decisions) Item 7 (problems)

In all three cases, values range from 1 to 5. Greater values are taken to indicate more communication problems.

Table 1 summarizes the variables of interest in the experiment and how the constructs have been operationalized.

Once the blueprint has been worked out and all the necessary provisions have been made, the experiment may begin. This is what happened:

8 Looking Behind the Curtain: The Experiment

The experiment spanned over five full days. It took place at the D-School on the Potsdam campus. The participants had to be present for the whole time, beginning from 9.30 each morning; on some days there were teams still working as late as midnight.

The project had been announced both as a "workshop on trauma" as well as an "experiment." It was made clear on all placards that the project was part of an experimental research program. Thus, the activities of participants would be observed and documented. At the same time, the program to be followed throughout the five days resembled that of a workshop. Participants would be supplied with information regarding trauma and had the task of developing some helpful approach.

40 students participated in the study, 15 men and 25 women. About half of the students had a technical background (software systems engineering). The background of the other students varied widely. Majors included business studies, languages, sports and others. On average, the participants were 22.71 years old and studied in the 4.82 semester. Half of the participants had been trained by the D-School, half of them not. We randomly assigned them to the mono- versus multidisciplinary team condition, making sure that there would be the same number of teams in each condition. Ideally, there should be three teams (of four members each) in all the four conditions:

1. D-School trained, multidisciplinary
2. D-School trained, monodisciplinary

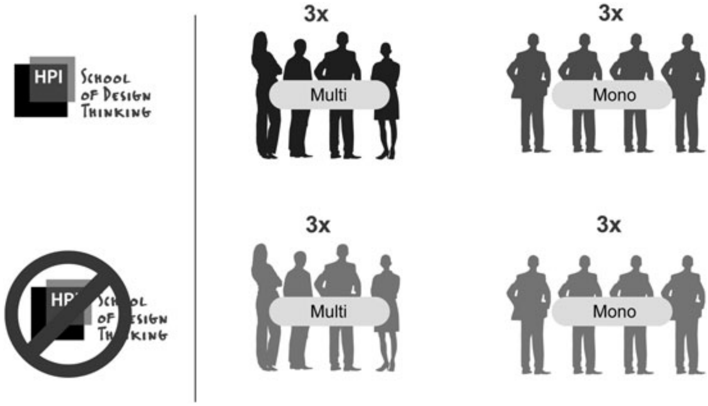


Fig. 1 The experimental setup allots three D-School trained multidisciplinary teams, three D-School trained monodisciplinary teams, three multidisciplinary teams without D-School training and three monodisciplinary teams without D-School training

3. Not-D-School trained, multidisciplinary
4. Not-D-School trained, monodisciplinary

Due to illnesses, there were some minor variations in the number of participants.

On each day of the experiment, multiple observations were made over and above those already specified. The participants filled out questionnaires regarding diverse issues such as their plan for proceeding, their satisfaction with their current standing, how they spent their time etc. A random sample of teams was filmed throughout the entire week, insofar as they were present at the D-School. Pictures were taken of all workspaces. The final presentations of all groups (approximately 10 min) were video-recorded. These video presentations as well as written summaries of the design solutions (1–2 pages) were made available online.

In the context of a lecture, the material was presented to trauma therapists and clients who had agreed to evaluate the solutions. The participants of the workshop/experiment were not present at that lecture so that personal sympathies or animosities would not bias the expert judgements (Fig. 1).

9 Design Thinkers Versus "Ordinary Students": Results

Of the two aspects of **innovation** that have been distinguished, lets consider **unusualness** first. D-School teams receive higher ratings than Non-D-School teams, as was hypothesized. The finding is consistent across experts and team members. Experts rate the unusualness of solutions by D-School teams with 2.80 on average; solutions by untrained teams 2.54. (Higher ratings indicate a greater degree of unusualness.) The participants themselves rate solutions by D-School teams 4.06 on average, solutions by other teams 3.65. The average unusualness ratings of experts

Table 2 Results regarding "usefulness" as estimated by the experts, comparing D-School trained teams with untrained teams

Question on usefulness	D-School	N	Mean	Mean diff.	p
What do you think, how helpful is this approach for the target group? (Experts, 1–5)	Trained	20	3.60	0.65	<0.05
	Untrained	24	4.25		
Which approaches should be realized absolutely? Please mark up to five approaches! (Experts, 0 or 1)	Trained	20	0.25	0.258	n.s.
	Untrained	24	0.42		

versus participants differ quite considerably in their height: Experts generally give lower ratings than participants. Thus, experts seem to have tapped the domain of potentially helpful interventions more completely than the project teams. Yet, the data consistently favors D-School teams in terms of unusualness.

Regarding the second facet of innovation, **usefulness**, all teams perform quite well. In none of the experimental conditions the average rating falls below "3," indicative of a "somewhat helpful" solution.

Just like the two measures of unusualness yield a consistent picture, the two measures of usefulness are consistent with one another too. However, the picture they suggest deviates from what had been expected. Not only does the data fail to show a significant superiority of D-School solutions. Indeed, Non-D-School teams outplay teams with D-School experience.

In Table 2, the column "N" specifies the number of ratings upon which the group averages are calculated. The column "p" specifies whether or not the difference between trained versus untrained teams is statistically significant. "N.s." means not significant, "<0.5" means significant and "<0.01" means highly significant.

Teams without D-School training receive higher ratings (4.25) on average than D-School trained teams (3.6). Higher values indicate a greater degree of usefulness; values may range between 1 and 5. The second measure of utility – whether or not a solution is chosen by the experts to be implemented "by all means" – points in the same direction. Solutions presented by teams without D-School training are selected more often (0.42) than solutions by D-School trained teams (0.25). Again, higher values indicate a greater utility; values may range between 0 and 1.

Now that we have considered trained versus untrained teams, lets take a look at the **mono-** versus **multidisciplinary** team condition.

Of all the groups, multidisciplinary D-School teams perform worst. Their average rating is close to 3 (somewhat helpful), whereas teams of all the other conditions receive an average rating above 4 (quite helpful) by the experts. Monodisciplinary teams outperform multidisciplinary teams, both in the D-School and in the Non-D-School condition.

Please note that statistical calculations for levels of significance depend not only on the size of the effect (here: the actual group difference) but also on the number of ratings. Thus, it is always a good idea to look at effect sizes over and above levels of significance. In Table 3, the average difference between mono- and multidisciplinary groups is greatest for D-School trained teams alone (first row in Table 3). It amounts to 1.083 as opposed to 0.167 for untrained teams (second row) or 0.633 for all teams

Table 3 Results regarding "usefulness" as estimated by the experts, comparing mono- versus multidisciplinary teams

	Teams	N	Mean	Mean diff.	p
D-School trained	Mono	8	4.25	1.083	0.05
	Multi	12	3.17		
Not D-School trained	Mono	12	4.33	0.167	n.s.
	Multi	12	4.17		
All teams	Mono	20	4.3	0.633	<0.05
	Multi	24	3.67		

together (third row). Yet, since the number of cases is halved when D-School teams are considered alone, the level of statistical significance is actually lower in the first row (for D-School teams only) than in the third row (where all the teams are considered).

Now, an interesting hook-up question may be whether there is some interrelation between unusualness and usefulness: Knowing that a solution is rather unusual (or usual), can you predict to some extent how useful the solution is? Or, vice versa, knowing that a solution is rather useful (or barely helpful), can you predict to some extent whether it is a rather unusual (or usual) solution?

Indeed, this is possible! The correlation between "unusualness" and "usefulness" is highly significant. It is negative: -0.547 ($p<.001$). This means, that the more unusual solutions are, the less they are helpful on average. (Correlations vary between -1 and 1. A value of zero indicates that there is no interrelation. A value of 1 indicates a perfect positive relation. A value of -1 indicates a perfect negative relation, that is: the higher the value of the first variable, the lower the value of the second and vice versa.) When only D-School teams are considered, the negative correlation between unusualness and usefulness becomes even more drastic: -0.700 ($p < 0.001$). This is an issue we will return to in the discussion.

Regarding **communication problems**, there is no statistically significant difference between mono- versus multidisciplinary teams; the effect sizes are negligible.

There is, however, a consistent difference between D-School trained teams versus untrained teams. According to all three indicators (items 5, 6 and 7), untrained teams experience more communication problems than teams with D-School training. This holds true both in the monodisciplinary as well as in the multidisciplinary team condition.

Teams without D-School training find it significantly more difficult to reach agreements (2.89 as opposed to 2.13). Members of not-trained teams report more group decisions they felt uncomfortable with (2.42 versus 1.88). Members of not-trained teams report more communication problems than members of D-School teams (2.53 as opposed to 1.88) (Table 4).

While some of the group differences fail to be statistically significant due to small N, it is noteworthy how consistent the picture is even when the mono- and multidisciplinary team condition are considered separately: All six comparisons indicate less communication problems in D-School teams (Table 5).

Table 4 Results regarding "communication problems", comparing D-School teams versus Non-D-School teams

Questions on communication problems	D-School	N	Mean	Mean diff.	p
Was it easy or difficult for your group to reach an agreement? (Item 5, teams, 1–5)	Trained	16	2.13	−0.77	<0.05
	Untrained	19	2.89		
Have there been group decisions that you felt uncomfortable with? (Item 6, teams, 1–5)	Trained	16	1.88	−0.546	n.s.
	Untrained	19	2.42		
Have there been communication problems in your team? (Item 7, teams, 1–5)	Trained	19	1.88	−0.651	<0.01
	Untrained	19	2.53		

Table 5 Results regarding "communication problems," comparing D-School teams with Non-D-School teams, multi- and monodisciplinary teams separately

		D-School	N	Mean	Mean diff.	p
Multi	Item 5	Trained	10	2.50	−0.600	n.s.
		Untrained	10	3.10		
	Item 6	Trained	10	2.00	−0.400	n.s.
		Untrained	10	2.40		
	Item 7	Trained	10	1.70	−1.00	<0.01
		Untrained	10	2.70		
Mono	Item 5	Trained	6	1.50	−1.167	<0.05
		Untrained	9	2.67		
	Item 6	Trained	6	1.67	−0.778	<0.05
		Untrained	9	2.44		
	Item 6	Trained	6	2.17	−0.167	n.s.
		Untrained	9	2.33		

10 Discussion

Regarding our two major experimental issues – **innovation** and **communication** – the second may be commented with greater ease as the findings approximate prior expectations. In terms of communication problems, no difference between mono- versus multidisciplinary teams has been found. Yet, D-School teams consistently report less difficulties than untrained teams. Does D-School training enhance communication skills so that communication obstacles may be handled more easily? Potentially. In pondering this causal claim, it needs to be considered that D-School trained team members generally knew each other in advance as they had studied together at the D-School. This familiarity yields an alternative explanation for reduced communication difficulties. Yet, quite a few of the untrained participants had known each other in advance as well. For example, most monodisciplinary teams comprised students of software systems engineering who knew each other from regular courses. Thus, there is some reason to assume that D-School training helps people to develop effective communication strategies. Whether the training does indeed have a causal effect in that regard, and what elements of the D-School experience most powerfully enhance communication skills, are issues that would have to be addressed by further studies.

More demanding, and potentially more interesting is the issue of **innovation**. Why were D-School teams, and multidisciplinary D-School teams in particular, outperformed by teams with no D-School experience?

A first reply might highlight the shortness of time available for the task. In a Design Thinking process, teams are encouraged to explore the problem space copiously before actually deciding on one particular solution. Indeed, this is what D-School teams did in the experiment. Untrained teams, on the other hand, were much quicker to decide. Quite a few of them selected their approach on the first day of the workshop. This left them with a lot more time for developing and refining a prototype. Following this line of thought, one might argue that D-School teams would have performed much better had they had a few more days to work on the project. Yet, this line of reasoning does not seem to endure careful consideration. After all, the experts did not rate the prototypes presented by the teams. These prototypes were, as a matter of fact, all rather foreshadowing than usable. What the experts did rate were the *ideas* teams had come up with. (If the suggestions were to be carried out, how helpful would they be?) D-School teams spent a lot of time selecting their idea, so the process of evaluation applied in the experiment should not work to their disadvantage. Thus, the supremacy of Non-D-School teams in our experiment calls for another explanation.

One important hint may be the strong negative correlation between **usefulness** and **unusualness**. Wild ideas are explicitly encouraged in the D-School training. While there is no need to question this outlook in general, there certainly is a danger of what may be called an **oddness trap**. When much effort is put into devising a solution that others will find surprising, solutions may be surpassed that are rather self-evident and yet highly effective. Indeed, these likely solutions may be the most effective ones in some circumstances. A "go-for-the-wild" approach might be more productive in circumstances when basically all likely solutions have already been explored and something else is wanted. In our experiment, this was obviously not the case. In all conditions, the average expert rating of "unusualness" falls between 2 and 3. That is, the experts state they have already considered the solutions presented by the teams, just not in all details precisely as the groups would have them.

In general, awareness of the oddness trap – knowing that there may be a trade-off between **unusualness** and **usefulness** – is only a first step. What we ought to strive for are means, strategies and potentially even techniques for avoiding the trap. Falling in love with funny ideas must not deflect designers from the user's true needs.

11 What We Wish to Pass Back

Having been endowed with a number of considerations by the Design Thinking community, we focused on a few recurrent believes. Now that the experimental results are in, our theory prototypes may be refined. In the dialogue between Design Thinkers and empirical reality, some hotspots have been identified that certainly span room for improvements. So, how can we sharpen our vocabulary? How can we refine our central believes so that they be ever more adapt to reality, ever more fruitful?

Regarding Design Thinking education, we might consider more explicitly what it is we wish to promote in differing circumstances. Certainly, there may be many situations in which fanciness or oddness is valuable in itself. In other cases, the users will want nothing but a working solution – whether fanciful or not. Maybe we can do a better job in systematising circumstances under which fanciness versus usefulness needs to be the ultimate standard. Maybe usefulness should always be the ultimate standard because fanciness trumps only when there is a major need for fanciness. In parallel to these theoretical issues, methodological considerations are likely as well: Should we equip students with (more) powerful methods to ensure a close(r) tie to the users' central needs? If so, ought we to provide a fixed procedure or would it suffice to make utility tests more explicit a factor in Design Thinking process models? Or, to name another possibility, should "carful utility tests" rather be taught as an overarching value/goal that students need to internalize?

Regarding the second experimental issue, we wish to turn to the advocates of multidisciplinarity in particular. Taking seriously the experimental results, some refinement in Design Thinking theory would seem helpful. This does not necessarily mean a major reorientation; some further specifications might due.

Perhaps multidisciplinarity does have a positive effect on innovation – but the effect is so small that it was easily overridden (and even "conversed") by chance variation in our experimental setting. If this is true, Design Thinking theory would surely profit from a realistic estimate of the effect size: If the effect size is small, we need to expect very limited gains with respect to innovation simply by assembling multidisciplinary instead of monodisciplinary teams. Or, to address another likely reasoning: Multidisciplinarity may have a considerable positive effect, but not in all contexts. For example, it comes to unfold its positive impact only after longer periods of time (months, not days). Another viable thought may be that multidisciplinary design teams provide more helpful prototypes than monodisciplinary ones when it comes to communicating design ideas to development divisions who work out final products. Such a handover was no subject of our experiment. Thus, there are many ways in which Design Thinking theory may be carried forwards by helpful specifications.

In sum, there is "experimental feedback" we may seek and use to refine Design Thinking theory – just as there is "user feedback" which design teams may seek and use to refine their prototypes. To be sure, this seeking and refining is a lot of hard work! And it may be a painful experience to see ones precious conceptions wobble under the pressure of an experimental test. But: We wouldnt be Design Thinkers if we were to duck out of the test, would we?

Innovation and Culture: Exploring the Work of Designers Across the Globe

Pamela Hinds* **and Joachim Lyon**

Abstract This chapter describes the preliminary results of a study of design practices in different regions and industries with the goal of understanding the relationship between culture, especially national culture, and the work of designers. In our ethnographic study, we have talked so far with 32 designers from Asia, Europe, and North American and observed designers as they did their work. We report initial insights about the role of the institutional context, especially client expectations, different attitudes toward what it means to be creative, different interaction norms within professions, different ways of using prototypes, and different ecologies around design education.

1 Introduction

Designers work every day in countries around the globe. They design cell phones, kitchen accessories, furniture, clothing, services, and just about everything we can imagine. How they do this and what it means to be a designer in different regions and cultures, however, is not well understood. In this research, we set out to understand and describe how design is practiced and what it means to be a designer in Asia, Europe, and the United States. Of course, this is a tall task and unlikely to be accomplished in a single year-long study. We did, however, gain insight into methods for doing such a comparative study and into a few key differences worthy of deeper exploration.

In this project, we take a broad, contextualized view of culture. Although we acknowledge that cultures are often characterized by different values (e.g. individualistic vs. collectivistic, see Hall 1963, Hofstede 1991), we focus more

P. Hinds (✉) and J. Lyon
Department of Management Science & Engineering, Stanford University, Stanford, CA 94305–4026, USA
e-mail: phinds@stanford.edu; jblyon@stanford.edu

* Principal Investigator

H. Plattner et al. (eds.), *Design Thinking: Understand – Improve – Apply*, Understanding Innovation, DOI 10.1007/978-3-642-13757-0_6,

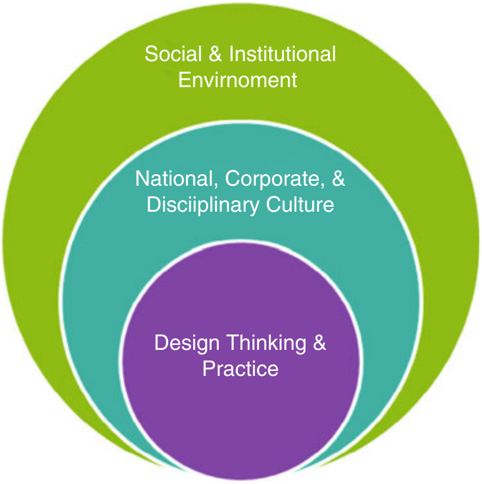


Fig. 1 A nested view of design thinking and practice

holistically on collaborative practices as situated in local and institutional contexts. Practices, the local context, the institutional context, and peoples' values, we believe, are intertwined and inseparable. In keeping with this view, we define culture as "a fuzzy set of attitudes, beliefs, behavioral norms, and basic assumptions and values that are shared by a group of people, and that influence each member's behavior and his/her interpretations of the 'meaning' of other people's behavior" (Spencer-Oatey 2000, 4).

We also subscribe to the point of view that culture is reflected not only in national culture, but is a mosaic composed of cultural identities derived from a variety of sources, including national culture, demographic features, and associations (see Chao and Moon 2005). In this study, we focused on national culture, but incorporated company culture and disciplinary culture as intertwined cultural forces that shape design practice. A nested view of culture considers the context in which people are embedded as instrumental in understanding behavior (see Perlow et al. 2004). Figure 1 reflects our perspective on how design thinking and practices are embedded within their cultural context. Each layer affects and is affected by the adjacent layers.

2 National Culture and Design Practice

In examining the related literature, we found no studies that explicitly investigated the relationship between national culture and design practice. One of the most relevant studies compared innovation management in Germany and China (Wang et al. 2005). They reported that German innovation activities were more clearly divided, sequential, and scheduled while the Chinese preferred to have overlapping activities. Khurana and Rosenthal (1998) also noted cultural differences in new product development, suggesting formality in US firms and a holistic approach in Japan. These studies support our *a priori* belief that design practices may vary by culture. A gap, however, remains in our understanding of how national culture is manifested in the actual *practice* of designers *in context*. Further, most studies to-date focus on cultural dimensions and not the social and institutional context in which people are embedded. In a review of economic development in China, for example, Zhao et al. (2006) reported that China may have difficulty incorporating the voice of the customer because of their Confucian value for stability over change and may adhere more strictly to supervision and rules. This could directly affect how design thinking is manifested in China. A recent article published in Interactions Magazine (Chavan et al. 2009) describes the limitations of design methods in emerging countries, suggesting that there is significant opportunity to evolve design practices to be more sensitive to those cultures. Chavan describes, for example, her Bollywood method that is more suited to the Indian market because it engages users in a dramatic Bollywood-style storyline as a means of transcending Indians' reluctance to give feedback in user studies. Our study extends previous research by asking how culture and the cultural context is embodied in innovation practices.

To explore questions about the relationship between culture and design practice, we are conducting ethnographic research. Ethnographic research rests on insights that emerge during the investigation. This approach enables us to understand the *meanings* that people associate with idea generation, prototyping, sketching, and other design practices, thus providing a deeper understanding of these perspectives than is available using quantitative methods.

3 Method

Our study involved interviews with and observations of designers as they worked. We have conducted interviews with designers and design managers in North America, Europe, and Asia. Each of these interviews generally lasted about 60 minutes with a range of about 30–90 minutes. Most interviews were conducted in private offices and meeting rooms, although sometimes the situation dictated that we conduct interviews in open spaces, cafes, and over meals. The interviews were semi-structured. Although we prepared an interview protocol, our primary goal was to understand the designers, how they worked, and what it meant to be a designer from their perspective, so the interviews were driven by what they told us was important and interesting. The interview protocol also evolved over the course of the study as we learned more and identified avenues for fruitful exploration. In general, we asked questions about a current or recent project and explored with them how that project was organized, the sources of innovation, how prototypes were used, and how that project reflected (or didn't) them as a designer. All interviews (unless the interviewee requested otherwise) were recorded and transcribed. Where possible, interviews were conducted in the native language of the designer either directly or, on occasion, through an interpreter. When interviews were conducted in languages other than English, they were transcribed and then translated by a professional translator or by the interviewer him or herself. In total, we conducted approximately 32 interviews of designers in China, 8 in Korea, 12 in Europe, and 12 in North America.

In addition, we reasoned that in order to understand the work practices of designers in different regions, we need to understand how designers were trained. We therefore conducted interviews with students and faculty in design programs in the U.S., Asia (Korea), and Europe (The Netherlands). We plan to conduct additional interviews with students and faculty in China, other programs in the U.S., and in other countries in Europe.

We also observed designers at several of the firms. Our preliminary work made it clear that observations of designers were necessary to compare the nuances of how practice differed between sites. During the summer of 2009, our research team spent approximately 7 weeks in China and Korea talking with and observing designers. We will observe at additional sites as the project continues. Our observations entail watching designers as they work and keeping detailed and extensive notes of their activities, conversations, frustrations, and successes. Thus far, we have observed for 7 weeks in China and Korea and for very limited amounts of time in North America and Europe. Table 1 captures our data collection strategy, including the type of data being collected in each type of firm and in each region.[1] Although we only indicate observations when we were able to spend an extended amount of time at the field site, in all cases, when we visited the site for interviews, we conducted ad hoc observations and include these field notes in our analysis.

Our goal in this study is to understand designers from multiple regions, disciplines, companies, and countries. Our sampling strategy was therefore to study design consultancies and inhouse design firms that spanned at least two continents and operated in different design spaces (see Table 1). With the help of the Zhejiang Innovation Center, we were able to negotiate access to two Chinese design centers, one of which has a European design operation. Although SinoCo did not have design centers outside of Asia, we included them in the study because they offered

Table 1 Sample by type of company and region

	North America	Asia	Europe
SinoFashion (inhouse)		17 interviews, 7 weeks of observations	Planned
SinoCo (inhouse)		7 interviews	
Elite (design consultancy)	12 interviews	13 interviews	12 interviews
Innovat (design consultancy)	Interviews and observations planned	9 interviews, observations planned	Interviews and observations planned

[1] To maintain the confidentiality of the firms being studied, we have not specified the exact countries in which the offices are located.

useful insights into home appliance design in China and had worked extensively with design consultancies inside and outside of China and were, as a result, able to share with us their perceptions of the client-consultant relationship. We were also fortunate to gain access to two design consultancies with global operations, one of which has invited us to conduct observations of their designers in North America, Asia, and Europe.

One of the challenges of conducting cross-cultural research such as this is that the data are richer and more complete if the interviews are conducted in the designers' native language and if the observers have adequate language skills to understand the conversations that are taking place in the work setting. We therefore composed a research team in which several members had Chinese or Korean language skills and were native to or had lived for an extended time in those countries. As we continue to collect data in Europe, we will augment the team as appropriate.

4 Insights

4.1 Culture and Design

Our preliminary research has revealed several insights related to culture and design practices including the role of the institutional context, especially client expectations, different attitudes toward what it means to be creative, different interaction norms within professions, different ways of using prototypes, and different ecologies around design education.

4.1.1 Client Expectations

One of our findings was that design practices can be strongly affected by the corporate culture, but adaptations also may be required to meet the demands of the local clients. The structures of local culture, institutional context (such as client demands), and organizational culture were intertwined and mutually determined the practices of designers. Client expectations, particularly at design consultancies, were described as heavily influencing the types of prototyping, concept generation, and storytelling processes followed by designers in different regions. Our data suggest, for example, that clients have *frames* that determine their expectations for what the designers should deliver, how they should behave, and of their own role in the process. These frames then affect the way in which design is interpreted and enacted. Specifically, we found the strongest effects from frames or expectations around building relationships, process as a deliverable, whether design activity is seen as a cost or an investment, and the value of form versus function. On most of these dimensions, the client environments in North America, Asia, and Europe were quite different.

Client-consultant relationships. Building and maintaining relationships with clients was universally seen as a critical aspect of being a designer in a design consultancy. What we found interesting though was that the nature of the relationship was experienced differently in different regions. In North America, for example, the relationship with the client was described as collaborative. Clients got involved in the activities of projects, including need finding, and this was seen as a way of "bringing the client along." In Europe, the distance between the client and the designers was described as greater with some resistance on the part of the client to "reverse" the client-consultant relationship or disrupt the power relationship by "being too friendly." One designer told us, "They are in charge and in the driver's seat, and they deliver that harsher than U.S. clients." Asia was similar, if not more extreme, in their concerns that the design consultancies were "reversing the service direction" by asking clients to make choices or approve of design concepts mid-stream. Asian clients were described as wanting the design consultancy to work on the project and deliver a final design based on their own counsel. In addition, contracts with Asian clients tended to be more fluid and there was an expectation of getting "extras" as a show of strength in the relationship.

Differences in these perspectives on the client-consultant relationship affected how design was done in each of these locations. When the project involved more participation of the client and both parties were perceived as having more or less equal status, such as in North America, designers anticipated regular feedback from the client which they would integrate into the design process on an ongoing basis. In these cases, designers saw their role as balancing client involvement in the process and, at the same time, maintaining the integrity of their designs. A critical practice involved managing the boundaries of client involvement so that they had adequate freedom within which to innovate. In Europe and Asia, the challenge for designers was to design in the absence of regular interaction with and feedback from the client, so although they had more freedom to design, they spent a larger percentage of their time working on design briefs to sell their ideas to clients at pivotal points in the project and generally seemed more anxious about client acceptance.

Process as deliverable. With regard to process as a legitimate deliverable, Asian clients were said to not care about how the designers approached the design activities. One informant told us that "They don't care [about] the process... just show me the cool stuff." In contrast, in North America, being a partner in the process was highly valued. Clients wanted to *own* part of the process. Different attitudes toward process as a deliverable affected design practices, for example, because Asian designers were required to produce concrete and visual forms of the design much more quickly while the North American designers could remain conceptual and even deliver representations of their *processes* (as opposed to the design itself) as legitimate forms of progress. European clients seemed to occupy the middle ground, although seeing what the product was going to look like was important. As one designer put it, the client says, "You're designers, show me visuals. I want to see what this is going to look like..."

4.1.2 What It Means to Be Creative

In our conversations with designers and design directors, we found significant differences between North America and Asia in what it means to be creative. We were surprised (and intrigued) in several parts of Asia to hear that Western designers were perceived as "too creative." When we probed for the meaning of this phrase, we learned that Western designers were sometimes seen as being radical and designing products that couldn't easily be manufactured or were only appropriate for a "niche" market – they were just too "out there." In Asia, particularly when observing inhouse designers, we noticed that using existing similar products as sources of inspiration was the norm. Day after day (in multiple environments), we observed designers sketching new designs as they looked back and forth from images of similar products in magazines or on related websites. When asked about this, designers in Asia told us that their goal was to design something that fit within the stream of existing products. Although preliminary at this point, we got the sense that while the North American and perhaps the European notion of creativity is to stretch beyond what is expected or known, Asian notions of creative design valued harmony with existing products. This finding resonates with cross-cultural research on individualism and collectivism which suggests that those from the West tend to prefer to stand out, to differentiate themselves from others, whereas those from Asian cultures tend to prefer to blend in and be seen as part of the group (see Nisbett 2004). Consistent with this, in design consultancies, we were frequently told that Asian clients would ask for a design that essentially replicated an existing "hot" product on the market. These requests were puzzling for Western designers, but made sense to designers who were born and educated in Asia.

4.1.3 Interaction Norms Across Professions

Although we found that, superficially, many of the same occupations and roles existed across regions, we are beginning to see interesting differences in the way that those occupations are constituted and the interaction among professionals from different occupations. In the U.S., for example, there is a strong value placed on multi-disciplinary teams within which designers of different stripes work closely together and even learn to make contributions that go beyond their own occupational boundaries. Mechanical engineers, for example, may get involved in user studies and anthropologists might make contributions that would generally be made by industrial designers. We saw some evidence that the lines between occupations were less blurred in Asia. In our observations of fashion designers in Asia, for example, we learned that there was less cross-training between designers and pattern cutters and that it would have violated the norms for designers to cut their own patterns. In France and England, we were told, fashion designers were trained to be pattern cutters and were expected to be able to take on this role, or at least demonstrate a high level of competency when working with a specialist pattern cutter. These occupational jurisdictions had a significant effect on what work designers did (and didn't do) in different regions and how they interacted with other professions as they implemented their design ideas.

4.1.4 The Role of the Prototype

Across regions, firms, inhouse vs. design consultancies, and industries, prototypes played highly social roles as objects around which questions were asked and social interaction occurred. In all cases, designers created prototypes throughout the design process and used these as a way of understanding their own designs and getting feedback from others. We found numerous objects referred to as prototypes, including models, sketches, scenarios, CAD drawings, garment patterns, etc. The creation and use of prototypes was somewhat determined by the disciplinary training (e.g. mechanical engineering, industrial design, fashion design, etc.) of the designer as well as the local context (e.g. client expectations, speed of the design cycles, etc.). Designers, not surprisingly, relied heavily on their disciplinary training in the type of prototypes they were likely to create and the types of questions they asked through the prototypes. Although this somewhat aligned with their role in the design process, it was not a perfect match. Mechanical engineers, for example, regardless of the stage of the design, were more likely to talk about CAD drawings and physical prototypes that enabled them to see how the mechanisms were going to work. Client (internal and external) expectations also determined when and how prototypes were created and used. External clients in Asia and Europe, for example, were more responsive to prototypes that were polished and resembled the "real thing" whereas North American clients tolerated prototypes that were strung together with duct tape and bailing wire.

4.1.5 The Ecology of Design Education

From the interviews that we have conducted with design educators and students, one early insight is that the ecologies around design education vary across regions. In all cases, there are multiple constituents including students, faculty, administrators, potential employers, government bodies, and professional associations. How these constituents interact with educational institutions, the amount of influence they have, and how these interactions shape design education and the future work done by students educated at these institutions varies by region. We are continuing to collect data and analyze the interviews and observations from design programs to explore these relationships.

4.2 Methodological Insights

Although the primary goal of this study is to understand how design practices vary based on the cultural context in which designers are embedded, we also gained insight into methods for doing this type of research and summarize those here. First, we begun the study with interviews and quickly realized that observations were necessary because interviews did not yield the level of detail required to compare the nuances of how practices differed between sites. We concluded that cross-cultural comparisons of design practices require observations of the designers as they engage with ideas, objects, each other, and their clients. Second, it became clear that the most insight was available when studying the design of products whose design was not "universal," e.g. the products are likely to be used/understood differently in different cultures, since these products are more influenced by the local cultures in which they were sold and used. When studying objects such as cell phones or laptop computers, there were significantly fewer differences than when studying products such as home appliances and fashion because the design of these objects was more intertwined with the culture and context in which they were being used. Home appliances, for example, need to account for the particular culinary preferences, available ingredients, and the configuration of kitchens in a given region. Third, consistent with research in cross-cultural psychology, we found that Asia, Europe, and North America were distinct regions. Although there are, of course, cultural differences within each region, even larger differences are in evidence between these three regions. As a result, including designers from North American, Asian, and European locations enables comparisons that facilitate broader understanding. Finally, we noticed significant differences in the design practices of inhouse designers vs. those working in design consultancies. Conclusions for one may not hold for the other. Sampling strategies, therefore, need to account for these differences either by systematically including and analyzing the different types of design activity or focusing on one or the other.

5 Conclusions

In this preliminary research, we have refined methods for studying design practices in different regions around the world and have gleaned insights about how and why practices might vary. This initial foray into research on the relationship between culture and design practices reinforces our idea that there are not universal "best practices" for design and that each region, country, and culture finds its own design path that leverages the culture and context in which the designers are embedded and is sensitive to the clients for whom they are designing.

We write this chapter with great caution. First, this work is very preliminary and our insights are the result of an, as yet, superficial analysis of the data. Although we have confidence that the insights that we write about in this chapter reflect what we were told by designers, more data collection and much more analysis is necessary to make sense of all that we learned and to derive deeper and more meaningful insights. Second cross-cultural research runs the risk of using and/or perpetuating cultural stereotypes. Our goal is to identify cultural differences and to build knowledge about and respect for the importance of these differences within the cultural context in which design work is being carried out.

References

Chavan, A., Gorney, D., Brabhu, B., & Arora, S. (2009). The Washing Machine that Ate My Sari – Mistakes in Cross-Cultural Design, *Interactions Magazine*, ACM.

Chao, G., & Moon, H. (2005) The Cultural Mosaic: A Metatheory for Understanding the Complexity of Culture. *Journal of Applied Psychology* 90(6), 1128–1140.

Hall, E. T. (1963). *The Silent Language*. Greenwich, CT: Fawcett Publications.

Hofstede, G. (1991). *Cultures and Organizations*. London: McGraw-Hill.

Khurana, A., & Rosenthal, S. R. (1998) Towards Holistic "Front Ends" In New Product Development. *Journal of Product Innovation Management* 15:57–74.

Nisbett, R. (2004). *Geography of Thought*. New York: Free Press.

Perlow, L., Gitell, J., & Katz, N. (2004). Contextualizing Patterns of Work Group Interaction: Toward a Nested Theory of Structuration. *Organization Science* 15(5), 520–536.

Spencer-Oatey, H. (2000). *Culturally Speaking: Managing Rapport through Talk across Cultures*. London: Continuum.

Wang, Y. Q., Cheng, Y. J., & Jiang, Y. Z. (2005). Cross cultural innovation management - a study of automotive technology system development. *Engineering Management Conference, 2005. Proceedings. 2005 IEEE International* Volume: 2, 760–764.

Zhao, X., Flynn, B. B., & Roth, A. V. (2006). Decision Sciences Research in China: A Critical Review and Research Agenda-Foundations and Overview. *Decision Sciences* 37, 451–496.

The Efficacy of Prototyping Under Time Constraints

Steven P. Dow and Scott R. Klemmer*

Abstract Iterative prototyping helps designers refine their ideas and discover previously unknown issues and opportunities. However, the time constraints of production schedules can discourage iteration in favor of realization. Is this trade-off prudent? This paper investigates if – under tight time constraints – iterating multiple times provides more benefit than a single iteration. A between-subjects study manipulates participants' ability to iterate on a design task. Participants in the iteration condition outperformed those in the non-iteration condition. Participants with prior experience with the task performed better. Notably, participants in the iteration condition without prior task experience performed as well as non-iterating participants with prior task experience.

Keywords Prototyping · Iteration · Empirical studies of design

1 Introduction

Many designers evangelize the value of prototyping [3, 7–9, 31, 37, 50], encapsulated in the design adage, "Enlightened trial and error outperforms the planning of flawless intellect." Prototyping entails repeatedly trying ideas and getting feedback [31]. A canonical prototyping iteration comprises four steps: envisioning possibilities, creating a prototype to embody a possibility, getting feedback about the prototype, and reevaluating constraints [29]. However, time constraints often lead organizations and individuals to focus on realization rather than iteration [3, 50].

S.P. Dow (✉) and S.R. Klemmer
Human-Computer Interaction Group, Stanford University, Gates Computer Science Building, 353 Serra Mall, Stanford, CA 94305, USA
e-mail: spdow@stanford.edu

* Principal Investigator

H. Plattner et al. (eds.), *Design Thinking: Understand – Improve – Apply*, Understanding Innovation, DOI 10.1007/978-3-642-13757-0_7,

This paper investigates if, under tight time constraints, several rapid prototypes yield more valuable design insights than allocating that time to a single iteration. Twenty-eight participants were randomly assigned to one of two conditions for an individual design task. Participants in the iteration condition were encouraged to test and refine their design multiple times. Participants in the non-iteration condition spent all their design time on construction; they were prevented from testing their design. After the design period, participants set aside all prototypes and entered a build period to implement their design.

The design task for this experiment was an *egg drop exercise* where participants design a vessel from everyday materials to protect a raw egg from a fall. This task has several appealing properties: success is objectively measurable (drop height), participants need only minimal technical expertise, there are many possible valid solutions, and it can be completed in an hour-long session. Drop height was the primary dependent variable. Participants also estimated their vessel's performance before and after the design period. We gathered participant demographics and concluded each session with a semi-structured interview.

The iteration condition significantly outperformed the non-iteration condition: the iterating participants' designs reached higher drop heights before breaking an egg. Self-assessment of performance increased significantly across the design period for individuals in the iteration condition. Unsurprisingly, participants with prior egg drop experience outperformed those without prior experience. More notably, non-experienced participants in the iteration condition did as well as experienced participants in the non-iteration condition.

Prior to describing our experiment, we summarize the existing literature that sheds light on the function and value of iterative prototyping.

1.1 Oscillating Between Creation and Feedback

Prototypes can help define an idea's role, implementation, and look and feel [26]; they can build empathy for users [8]; they communicate to clients, users, and fellow designers [56]. Designers embody creative hypotheses in prototypes and then observe the outcome [31]. An iterative prototyping practice oscillates between creation and feedback: creative hypotheses lead to prototypes, leading to open questions, leading to observations of failures, leading to new ideas, and so on.

In the creation phase, designers ask the abductive question of "what might be" [43, 44]. Much of previous design research has emphasized the importance of creative idea generation [6, 34, 45, 47, 55]. Research on brainstorming [6, 45, 47, 51], synthesis [34], and framing [22, 57] techniques seeks to improve the abductive part of prototyping. Expertise literature suggests expert practitioners develop an organizational framework for retrieval and application of knowledge [17]; expert designers learn to effectively organize and act on locally contextual design information.

In the feedback phase, designers make inferences from observations [35]. Experimentation and feedback leads designers to discover unknown attributes, constraints, and opportunities that may not have been conceived of a priori. Discovery is not an automatic consequence of experimentation; the way people frame problems makes some insights salient and hides others [32].

1.2 Prototyping with Internal and External Representations

Designers can use mental imagery to envision and improve ideas [2, 18–20]. Christensen and Schun analyzed an engineering design setting where designers use mental simulation as a proxy for external prototyping, reducing "uncertainty language" within meetings [4, 10]. Similarly, Schön remarked that an expert designer possesses the ability to conduct a series of "what-if" moves with "discovered consequences, implications, appreciations, and further moves" [48]. But as Schön points out, the web of moves can become too complicated to manage in one's head – even for virtuosos – due to limitations in human memory and processing.

People leverage the physical world to overcome limitations in memory capacity [5, 46], to convert highly cognitive tasks into perceptual/motor tasks [12, 25, 27, 38], to effectively represent problems [36, 58], and to explore alternatives [33, 41, 42]. Kirsh and Maglio's study of the game Tetris found that players manipulated the pieces more than was pragmatically necessary for moving them to the right place [33, 42]. Kirsh and Maglio argue that these manipulations provide an epistemic technique for exploring alternatives. Prototypes are designer's way of trying things out.

Larkin and Simon [36] explored the representational differences between a diagram and a written description. They demonstrate two external representations may be informationally equivalent, but have significantly different computational efficiency. Designers' choice of external representations in prototyping has significant influence on how they explore a design space [9, 21, 39, 40].

Tversky and Suwa investigated how external representations promote discovery and inference. They show that by attending to visual features in sketches, designers discover ideas that were unintended when they were drawn [52, 53]. Prototypes similarly elicit information about the design context that did not previously exist in the designer's head.

1.3 Is Iterative Prototyping Undervalued?

Design is often heavily time-constrained; this can discourage designers from iterating. Many feel that organizations undervalue iteration [3, 16, 30, 49, 50]. Prototyping has an actual bottom-line cost associated with it, but this cost estimate is often inaccurate or changes over time [3]. Organizations often avoid prototyping because they believe the cost/investment will be significant and the return will be minimal. As Schrage suggests, "it is hard to persuade companies that one more iteration costs less than a flawed product," [50]. While researchers have devised economic models and performed cost-benefit analysis to argue for rapid iteration [16, 30], resource considerations remain a primary barrier to its application in industry.

On the view of prototyping as a learning process, psychological explanations of learning barriers can provide insight into why prototyping may happen too little in practice. Dweck has demonstrated that people's belief in whether intelligence is mostly fixed or mostly shaped by practice has a significant impact on whether people seek out learning opportunities [14]. Dodgson and Wood have shown that with high self-esteem, people respond less negatively to failure and focus on strengths rather than weaknesses [13]. Earnest experimentation requires risk. The educational psychology literature can inform how to structure the environment so that designers fully engage the prototyping process [1, 14].

2 Method

The design task had two conditions: individuals encouraged to conduct iterative testing (iteration) and individuals prevented from conducting iterative testing (non-iteration). We tested the following hypotheses:

- Participants in the iteration condition will outperform the non-iteration group.
- Participants in the iteration condition will report a larger increase in pre/post confidence levels (perceived ability) than the non-iteration condition.
- Participants with prior exposure to the design task will outperform participants with no exposure.
- Participants with prior general design experience will outperform participants with no design experience.

2.1 Materials and Design Task

In selecting the experimental task, we sought to achieve the following four criteria:

- Presents a clear, objective measure of design quality
- Requires minimal design or engineering expertise
- Can be completed by individuals within one hour
- Offers many paths to achieve an effective result.

We chose the *egg drop exercise*, where participants design a vessel from everyday materials to protect a raw egg from a fall. Variations of the exercise are practiced in secondary and tertiary education classrooms around the United States. This study measures performance by dropping a single egg from a one-foot marker, then two, then three, and so on until the egg cracks. Task performance is measured by the highest height (in feet) at which the egg survives a fall.

Fig. 1 *Left*: Materials constraints in the design task: pipe cleaners, popsicle sticks, rubber bands, tissue paper, poster board, and flat foam. *Right*: Experimental setup for the design exercise

Pilot studies showed that our choice of materials should be diverse enough to elicit many approaches yet challenging enough to produce a wide range of performances. We selected the following design materials: eight pipe cleaners, eight rubber bands, eight popsicle sticks, one $4'' \times 8''$ piece of poster board, one sheet of tissue paper, one $4'' \times 6''$ piece of flat foam, and 1 ft of scotch tape (Fig. 1, left). Participants worked on a table next to a drop zone area with foot markers written on the wall (Fig. 1, right). All of the supplies were on the table, including build materials, scissors, eggs, and instructions.

For their participation, subjects received either credit towards their course research participation requirement or a $20 Amazon gift card. As additional incentive, participants were told the two best performing vessels would receive additional Amazon gift cards.

2.2 Participants

Twenty-eight students averaging 21.1 years old and representing a wide range of majors from our university participated in the study. Participants were randomly assigned to one of two conditions. The study balanced for gender, prior egg drop experience, and general design experience across the two conditions. Twelve of the participants had prior experience with the egg drop exercise. Six had either worked as product designers or participated in regular design activities.

2.3 Procedure

Participants filled out a consent form and demographics questionnaire. The experimenter verbally described the egg drop exercise and the specific rules for the assigned condition. All participants were told they would have 25 min to **design**.

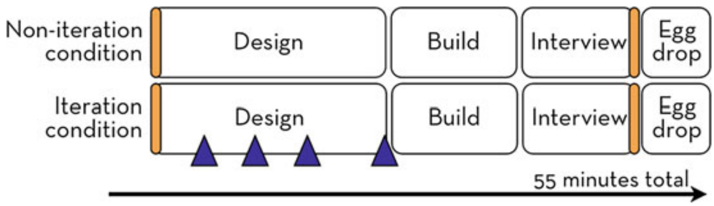


Fig. 2 Experiment procedure with time markers for requesting tests in the iteration condition (*triangles*) and for requesting task performance estimates (*vertical bars*)

They were given a set of construction materials, and were told they could get replacement materials if necessary. After the design period, the researcher cleared the workspace and provided a fresh set of the original materials (this time without replacements). Participants were given 15 min to **build** the final design, followed by a 10-min **interview**, and the egg drop **test** (Fig. 2).

During the design period, participants in the control group (no iteration) were provided one egg, which was also used in the final egg drop. Individuals in the manipulation group (iteration) were given a full carton of eggs. We encouraged iteration participants to conduct a test drop at the 5, 10, 15, and 25-min marks during the design phase. We did not limit participants to only four drops, nor did we strictly enforce all four drops. The drop zone was adjacent to the design table so participants in the iteration group could test their design ideas at any point (Fig. 1, right).

Participants were asked to estimate their perceived performance on the task (in feet), both after hearing the instructions and right before the egg drop test. We conducted a short open-ended interview at the end of the build phase, asking participants to describe their concept and their biggest concern for how the egg might break.

3 Results

This section describes the effect of iterative testing on task performance, the effect of iterative testing on task confidence, and the influence of prior task exposure on design performance.

Vessels created in the iteration condition outperformed the non-iteration condition, with an average successful egg drop height of 6.1 ft compared to an average of 3.3 ft ($t = 2.38,\ p < 0.03$) (Fig. 3).

Participants' confidence level in the iteration condition rose from an average of 4.14 to 5.93 ft from before to after the design task ($t = 2.21,\ p < 0.05$). The non-iteration condition saw no significant change in perceived ability, averaging 3.1 for both pre and post design task (Fig. 4). The pre-measure of performance slightly favors the iteration condition, although the mean self-estimates are not significantly different ($t = 1.92,\ p = 0.23$).

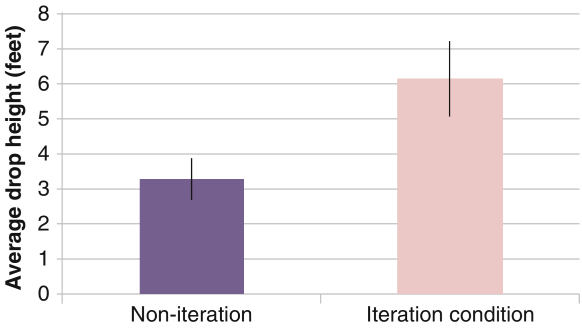
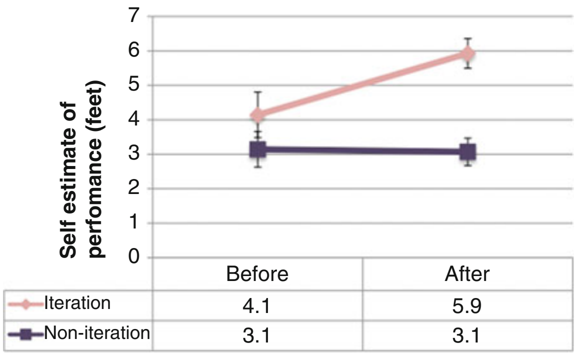


Fig. 3 Individuals in the iteration condition significantly outperformed the non-iteration condition in the egg drop mechanical design task

Fig. 4 Individuals' self estimate of performance (measured in feet) – shows a significant rise between pre- and post-task estimate, but only in the iteration condition



Participants in both conditions estimated their performance fairly accurately. On average, iterators estimated 5.9 ft, just underestimating their actual score of 6.1 ft, and non-iterators estimated 3.1 ft, just underestimating their actual score of 3.3 ft.

3.1 Influence of Prior Exposure to Design Tasks

Twelve of the twenty-eight participants reported previously taking part in the egg drop exercise. Prior egg droppers outperformed those without experience, 6.3 ft compared to 3.5 ft ($t = 1.98,\ p < 0.04$) (Fig. 5).

Both experienced and inexperienced participants in the iteration condition outperformed their counterparts in the non-iteration condition (Fig. 6). A two-way repeated measures analysis of variance (ANOVA) was performed, with Iteration (iteration/non-iteration) and Prior Experience (prior/no-prior) as factors and egg drop height as dependent variable. Participants with prior egg drop exposure in the iteration condition performed the best, with an average successful drop height of 8.7 ft compared to 3.8 ft for prior egg droppers in the non-iteration condition ($F(1,26) = 6.84,\ p = 0.015$). Similarly for participants with no prior egg drop

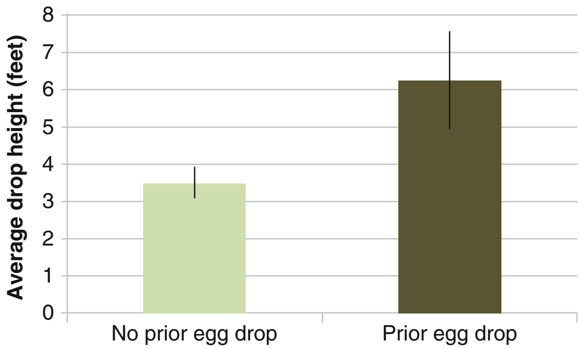
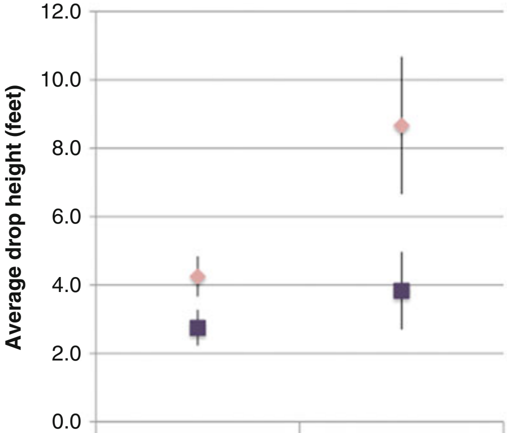


Fig. 5 Individuals with prior exposure to the egg drop task significantly outperformed those who had not done this exercise before



	No prior egg drop	Prior egg drop
Iteration	4.3	8.7
Non-iteration	2.8	3.8

Fig. 6 Breakdown of participants with or without prior egg drop exposure and those in the iteration or non-iteration condition (chart and table numbers in feet)

exposure, the iterative testing condition outperformed the non-iteration condition, 4.3 ft compared to 2.8 ft ($F(1,26) = 5.93,\ p = 0.023$). The Iteration x Prior Experience interaction was nearly significant ($F(1,26) = 2.45,\ p = 0.130$). Iteration helped participants with no prior egg drop experience perform at the same level as non-iterators with prior egg drop exposure.

3.2 *Influence of Design Experience on Task Performance*

Six of the twenty-eight participants had prior professional product design experience or participated in regular design activities. Prior design experience had no significant effect on the outcome of design task performance ($t = 1.84,\ p < 0.17$). With only six qualifying participants, the sample size is not large enough to fully explore the effects of prior design experience.

4 Participant Creations

Participants explored a wide variety of creative design concepts including parachutes, damping stilts, tubes, boxes, suspension systems, and nests for catching the egg raw (Fig. 7). The top three performers – 15, 13, 10 ft came from the iteration condition (top of left column). Based on these participant creations, we conducted an analysis of the design space [15] and determined five key design dimensions: the amount of drag created in the air, the distance between the egg and the first point of impact, the damping upon impact, the balance of weight before and after impact, and the containment of the egg. While this analysis of the design space is informal, it sheds light on relevant design factors. The interviews provide further insight on how participants discovered important variables, and typically focused only just one or two of these factors.

5 Interviews

The interviews revealed how participants employed different prototyping strategies, learned from iteration, and used mental simulation.

5.1 Prototyping Strategies

Some participants employed their understanding of physics to build a vessel designed to absorb impact. For P22's vessel (Fig. 8, left), he coiled "the foam (into) a spring to absorb the shock." P24 said she included a "stabilizing layer" for "bigger surface area" and so "the force was a little more dispersed." As P3 explained "the part that hit the ground had the most impact, so I didn't want that part to be the egg." Her design, a self-described "spiky creation," included damping stilts protruding in many directions. According to another participant, P18, the key was to provide a "buffer," so the "impact point doesn't hit the egg directly."

Other participants approached the egg design task as bricoleurs. As P23 described, "I started with a poster board box and then lined it with the foam box, and then I tore up little pieces of foam 'cause I had extra. And then, 'cause I had 'em I threw in the pipe cleaners around the top of the egg. On the bottom there are sticks, partly because I had 'em, but also it makes it more likely to land on the bottom." P25 (Fig. 8, middle) simply wrapped the egg with as many layers of materials as possible. This approach of mashing together materials echoes the opportunistic design practices reported by Hartmann et al. [23].

Other participants drew inspiration from objects outside of the immediate design context. P11 said, "My design is a Turkish cone. This is the same thing they use to sell chestnuts… When I drop chestnuts usually they would not crack, although

Fig. 7 Twenty-eight participant creations ordered according to best performers (from top left down) and separated by study condition (iteration in left two columns; non-iteration in right two columns)

[chestnuts] are much harder" than eggs (Fig. 8, right). In a similar vein, P21 related the design task to protecting passengers in vehicles. Both participants thought of analogous situations for protecting precious objects.

Fig. 8 *Left*: Participant 22 (iteration) successfully protected the egg at 13 ft. *Middle*: Participant 25 (non-iteration) protected at 7 ft. *Right*: Participant 11 (iteration) protected at 5 ft

Fig. 9 *Left*: Participant 3 (iteration) successfully protected the egg at 5 ft. *Middle*: Participant 18 (iteration) protected at 2 ft. *Right*: Participant 28 (iteration) protected the egg only at 1 ft

Many participants used the materials and gestures to communicate about the features of their vessel. P20 said "I designed an outer boundary [hands around the prototype], using the [looks up at reference sheet] pipe cleaners… and I designed an inner boundary using the [look up at sheet] sticks."

5.2 Learning from Iteration

Most participants in the iteration condition made a concerted effort to learn and improve their designs with each iterative test. As P3 stated, "experimentation with materials is important, especially at the beginning, so you figure everything possible you can do with them. It is also really important to see what actually happens when it hits, 'cause with my first design I didn't realize it would hit so hard." Her main design insight was to have damping sticks protruding at different angles (Fig. 9, left).

P9 learned the vessels do not fall evenly: "What I didn't account for is, as it gets to higher heights, this will not drop straight down." P18 recognized a different problem with her design. "The main problem with the last design is that it wasn't covering the egg enough, so I was afraid it was going to fall out" (Fig. 9, middle). The iterative process helped participants identify issues such as creating drag, balancing the weight, managing the landing, and containing the egg.

Iterative testing does not always reveal the source of failure. P28's first vessel braced the egg with a square wooden structure. It broke at 1 foot. Then he added a platform underneath and parachute, while he kept the wooden frame (Fig. 9, right). Although, his design continued to fail from low heights, P28 never inferred a key problem: the wooden frame can easily jab into the egg, cracking it with very little force. However in the final interview he did say "I was thinking that I could use – for the holding bay – instead of the wood, I could actually use the pipe cleaners since they are a little bit softer."

5.3 *Using Mental Simulation*

The interview right before the final egg drop asked participants to envision how their concepts performed. P12 commented and gestured using his design, "When it falls, it's probably going to fall one way or another. Once it starts getting dropped from higher, it's going to bounce and flip maybe [shows how the vessel might flip over]." P14 projected his design would "land kinda crooked sideways." P27 was concerned her design would impact the floor on its side (Fig. 10). She also correctly observed that the parachute should keep her design from falling on its side.

5.4 *Effects of Manipulating Iteration*

In the iteration condition, many participants expressed frustration with having to drop so early and often. At one point during the task P16 says to himself: "What is sturdy enough to support an egg drop?" Then he sighs, takes a deep breath, and sits back in his chair looking frustrated. He felt pressured to come up with something under the tight time constraint. In the interview he said, "I thought five minutes was too soon to really have anything substantial" (Fig. 11, left). While the tight iteration cycles were stressful, his vessel scored 6 ft – average for the iteration condition

Fig. 10 Participant using her vessel to illustrate possible failure scenarios

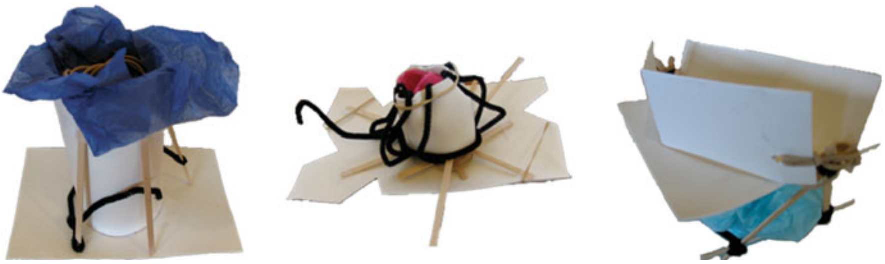


Fig. 11 *Left*: Participant 16 (iteration) successfully protected the egg at 6 ft. *Middle*: Participant 21 (iteration) protected at 7 ft. *Right*: Participant 15 (non-iteration) protected the egg at 2 ft

and significantly better than the average non-iteration score. Other participants embraced the opportunity for iteration and really stress-tested their vessels, such as P22 who stood on the table to test his design.

In the non-iteration condition, participants were often ready to test their idea before the end of time period, as P15 said, "so if I'm done can I start?" Similarly, with 10 minutes left in the design period P25 declares, "Alright, I'm finished." It was not clear (at least to us) a priori that the multiple-iteration condition would be so much more engaging for participants than the single-iteration condition.

5.5 *Iteration did not Lead to Divergence*

While participants in the iteration condition were allowed to test multiple egg drops, they did not necessarily explore a variety of concepts. As P16 described, "I'm not a very good outside-the-box thinker, so I kinda just had one idea and I was going to try to make it work." P27, who had the best overall design, expressed a similar notion: "I went with the whole parachute idea…from the beginning. So, I had one core idea." Generally participants selected an initial design direction and iterated to improve on that idea.

More unexpectedly, some participants claimed that their chosen design seemed like the only possibility. P21 said, "For some reason this seems to be the only idea. There needs to be a platform and then as good of cushion as possible. I don't see any other way" (Fig. 11, middle). Likewise P20 asserted, "This is the best approach for such a design." Despite oft-mediocre preliminary tests and a wide range of possibilities available, many participants appeared fixated on their initial design concept.

5.6 *Factors that Prevented Divergence*

The short time period impacted why participants did not diverge. As P18 stated, "This is what I thought of first [holding his design], and I started thinking, 'well that's one idea what else can I do?' Then I said, 'nah, I better make this to make sure I will have time." P24 discussed the notion of changing to a new idea, "With time and with trials, I was sort of improving upon the first idea I had and not trying to scrap it and go on to a whole new idea." Participants may not have felt they had time to brainstorm different ideas, and once they got started, they found it difficult to justify changing to a new idea.

While many participants described how they had "one idea and just went with it" (P6), some participants indicated ideation occurred before prototyping. P27 talked about constructing "some sort of box with the sticks and involving rubber bands so the egg is in the middle." P24 said, "I think if I had more time I probably would have been more accurate, maybe even do some calculations." Participants may have considered ideas that were not pursued due to lack of time and perceived complexity. P4 commented: "There were a lot of different ideas I had originally… possibly even using the tissue paper like a parachute."

Participants' underlying assumptions affected their fundamental design choices. P15, like others in the experiment, assumed the egg had to drop by itself into a nest: "I just figured I was supposed to build a vessel to catch the egg on its own" (Fig. 11, right). P11's sense of personal pride in his "Turkish cone" perhaps dissuaded his willingness to pursue other concepts between iterations: "An [alternate] design may have been better… but I am proud of mine." (P11)

6 Conclusion

Participants entered the final fifteen-minute build period armed only with what they learned during the design period. Why did participants in the iteration condition outperform non-iterators? One interpretation says that participants in the iteration condition discovered more flaws and constraints, and tried more new concepts. Non-iterating participants could only speculate how their design would perform. Another interpretation says participants in the iteration condition became better carpenters; they often built the same construction multiple times and thus they tuned the craft. These interpretations are not mutually exclusive, as experimenting and discovering constraints are part of craftwork.

Why did participants in the iteration condition significantly increase their estimated performance on the design task? Unlike the non-iteration condition, the iterating participants received multiple benchmarks. Each iterative test contributed to their judgment of performance. Participants in the non-iteration condition also managed to correctly estimate their low performances, so it remains inconclusive whether the feedback alone leads to better self-estimates. Surprisingly, the non-iteration condition saw no rise in perceived performance despite working on the task for 40 min.

Why did iterating participants with prior experience far outperform all others? Prior exposure to the egg drop exercise gave participants a head start in forming initial design concepts, but why did they make stronger gains with feedback than preliminary ideas from newbies? One argument says that prior experience gives people an index of examples (or cases) and feedback merely aids people to sort through the good and bad ideas. Another argument says prior experience is not only about knowing examples; it's about knowing how to perceive and analyze feedback on proposed solutions. This finding suggests the possibility for scaffolding design expertise with domain-specific examples, along with various feedback perspectives. Future authoring tools, for example, could include domain-specific design exemplars, each with a host of expert feedback.

What factors influence the use of rapid iteration? We found some participants expressed anxiety from having to iterate too early and too frequently. The iteration condition demanded proficiency and imprecision. On the other hand, several of the non-iterating participants were unsatisfied because they could not immediately see how their design performed. Participants may favor longer iterations over short and early iterations to avoid duress; this emotional factor may affect design outcome.

Iterative prototyping does not necessarily lead to an exhaustive exploration of alternatives. Participants in both conditions of the study explored a narrow range of possibilities in the design space. The short time frame and uncertainty about more complex constructions influenced participants. Unlike many real-world design processes, the design period did not include structured time for divergent thinking. More interestingly, several of the participants talked about how they believed their idea to be the only possibility. Design research explains people often fixate on concepts, especially if they have invested energy and time into one path [11, 28].

External validity is a concern for any lab study. While most real world design ventures are often social in nature, we focused on individual designers in this preliminary study to avoid the potential confounds of groupthink and interpersonal relationships. Likewise, design problems are typically solved over the course of days or months. To control for external stimuli, we chose a time frame that only required a single uninterrupted session. Our choice of a design task placed value on having an objectively measurable outcome. In the real world, the problem space or "design brief" is often not set in stone; it gets defined along the way. That said, the egg drop design exercise might be in some respects representative of design tasks that do have clear goals (e.g., designing a bridge always has a clear objective: to insure that cargo and people can cross safely). As a whole, participants demonstrated a range of creative solutions to the egg drop problem. Just as in real design settings, the outcomes cannot be defined by success/failure/right/wrong, but by what concept best fits the current design context.

7 Future Work

Questions remain about how designers perceive the efficacy of prototyping. Do designers undervalue rapid iteration? Within a given timeframe, how do designers determine an iteration strategy? How do designers decide the frequency and temporal spacing of iterations? Do designers typically plan iterations or do they unfold