First International Conference on Design Creativity, ICDC 2010
29 November - 1 December 2010, Kobe, Japan

"Mind the Oddness Trap!" - Theory and Practice in Design Thinking

Christine Noweski and Christoph Meinel
Hasso-Plattner-Institute for IT-Systems Engineering GmbH, Germany

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. It should help to test and refine theory and 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 (e.g. at D-Schools or in art curricula) produce more innovative solutions than untrained teams. Thus, hotspots have been identified that may stimulate some productive refinements of design thinking theory.

Keywords: design thinking, interdisciplinarity, innovation, creativity, teamwork, creative space empirical design research

1 From Design Thinking to Design Thinking Research

In recent times, the term 'design thinking' has received attention in various fields of interest. The concept has its roots in research on how designers comprise wicked problems and develop novel and viable solutions. Originally investigated in domains like architecture and industrial design, the initial research focuses on cognitive models supporting the generation, condensation, and creative transformation of design knowledge and design concepts (Cross 1990; Lawson 2006; Rowe 1991). Building on that, design thinking was further developed and translated into metadisciplinary frameworks (Lindberg et al. 2010) detached from designers' professional domains and was applied to various disciplines and fields of innovation. Design agencies such as IDEO promote working methods labelled with this term and inspire large scale companies like Procter & Gamble and SAP to 'design thinking' approaches to innovation (Brown 2008; Holloway 2009; Martin 2009). Also, the term has expanded into academic curricula beyond traditional design programs, as, for instance, at Rotman School of Management (Toronto) in the context of MBA education, and at the d.Schools in Stanford and Potsdam which offer design thinking education specifically to non-designers (Dunn & Martin 2006; Plattner et. al. 2009).

In face of this broad range of 'design thinking' usage, we believe it to be essential to analyze sharply, what really matters for it to be successful. To do so, one needs to confront the complex concrete world. 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.

2 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 believes in the field. Our take was this: In most literature resources, it is assumed that design thinking fosters innovation. From this, we conduct the hypotheses, that people who have been trained in design thinking should produce more innovative solutions than people who have not been thus trained. There are multiple institutes who offer design thinking education. We will name this kind of education 'D-School training' in the following. Our starting hypothesis may thus be formulated more specifically: It is assumed that D-School trained teams produce more innovative solutions than teams without such training.

Additionally, to consider one rather confined factor, we shall test the widespread belief that interdisciplinarity enhances innovation. If the belief is correct, interdisciplinary teams produce more innovative solutions than monodisciplinary ones on average.

While the two hypotheses concerning D-School training and interdisciplinarity 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. We therefore, broke the abstract concept down into disparate factors that we then could assess more easily. This is our take: *A design solution S1 is considered more innovative than a solution S2 if S1 is more unusual as well as more useful than S2.*

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:

D-School trained teams produce more unusual solutions than teams without this training.

D-School trained teams produce more useful solutions than teams without this training.

Accordingly, two hypotheses may be formulated concerning interdisciplinarity:

Multidisciplinary teams produce more unusual solutions than monodisciplinary teams.

Multidisciplinary teams produce more useful solutions than monodisciplinary teams.

In interdisciplinary teams, 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 (reconsider 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 interdisciplinary teams produce more unusual results than monodisciplinary teams.

Regarding the second facet of innovation – usefulness – the development of such solutions depends upon knowledge, e. g., knowledge concerning the situation of users or knowledge about technical options for realising 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, interdisciplinary teams seem better equipped for developing useful solutions.

Yet, at the same time, there is a reason to believe that, on average, interdisciplinary teams will produce less innovative solutions than monodisciplinary ones. Why that? Even if interdisciplinary teams have a greater potential for innovation, communication problems might hinder them. It seems reasonable to expect that communication will be more challenging in interdisciplinary 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 categorisation 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:

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 interdisciplinarily. 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:

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

3 Operationalization or: Let’s Get Concrete!

Now that a challenge has been specified the question of how to asses the attributes of interest needs to be considered. We opted for an expert rating. Each working team was given the same challenge (to develop a solution to facilitate the everydaylife of traumatized patients), the same equipment, space and time and then, had to present its solutions to the other teams and professionals of the specific medical fields. But still, we had to handle he unusualness of each solution to gain reliable measures result. How is one to count the unusualness of a design solution, for instance? Obviously, some further steps needed to be taken.

In order to assess abstract factors they need to be operationalised. 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 was to find concrete entities that one could look at to arrive at reasonable statements about the abstract notions of interest.

At level of theory there were 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) to (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.

4 Looking Behind the Curtain: The Experiment

The experiment spanned over five full days. It took place at the D-School spaces at Potsdam University campus. The participants had to be present for 5 days, beginning at 9.30 each morning untill a self choses end - some teams were 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 announcements 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 interdisciplinary team condition, making sure that there would be the same number of teams in each condition. Ideally, there should be 3 teams (of 4 members each) in all the four conditions:

1. D-School trained, interdisciplinary,
2. D-School trained, monodisciplinary,
3. Not-D-School trained, interdisciplinary,
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 stand- ing, 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 minutes) 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 work- shop/experiment were not present at that lecture so that personal sympathies or animosities would not bias the expert judgements.

5 Design Thinkers vs. "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 un-usualness.) 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 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 favours 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.

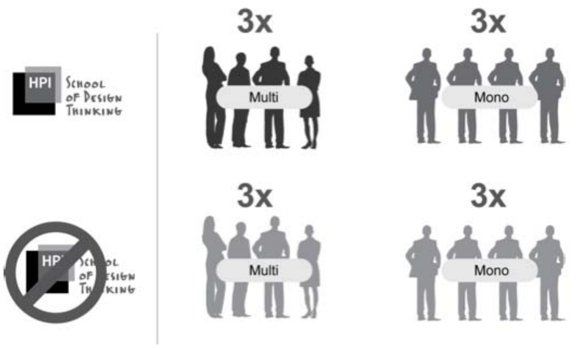

Fig. 1 The experimental setup allots 3 D-School trained multidisciplinary teams, 3 D-School trained monodisciplinary teams, 3 multidisciplinary teams without D-School training and 3 monodisciplinary teams without D-School training.

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.

Table 1 Results regarding "usefulness" as estimated by the experts, comparing D-School trained teams with untrained.

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	.65	<.05
	untrained	24	4.25		
Which approaches should be realized absolutely? Please mark up to five approaches! (Experts, 0 or 1)	trained	20	.25	.258	n.s.
	untrained	24	.42		

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 "<.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 "absolutely" – points in the same direction. Solutions presented by teams without D-School training are selected more often (.42) than solutions by D-School trained teams (.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 interdisciplinary team condition. Of all the groups, interdisciplinary 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 interdisciplinary 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 interdisciplinary 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 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: -.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: -.700 (p<.001). This is an issue we will return to in the discussion.

Regarding communication problems, there is no statistically significant difference between mono- versus interdisciplinary 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 interdisciplinary team condition.

Table 2 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	.05
	multi	12	3.17		
Not D-School trained	mono	12	4.33	.167	n.s.
	multi	12	4.17		
All teams	mono	20	4.3	.633	<.05
	multi	24	3.67		

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 3 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	-.77	<.05
	untrained	19	2.89		
Have there been group decisions that you felt un-comfortable with? (Item 6, teams, 1-5)	trained	16	1.88	-.546	n.s.
	untrained	19	2.42		
Have there been communication problems in your team? (Item 7, teams, 1-5)	trained	19	2.53	-.651	<.01
	untrained	19	2.53		

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 multi- disciplinary team condition are considered separately: All six comparisons indicate less communication problems in D-School teams.

Table 4 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	-.600	n.s.
		untrained	10	3.10		
	Item 6	trained	10	2.00	-.400	n.s.
		untrained	10	2.40		
	Item 7	trained	10	1.70	-1.00	<.01
		untrained	10	2.70		
Mono	Item 5	trained	6	1.50	-1.167	<.05
		untrained	9	2.67		
	Item 6	trained	6	1.67	-.778	<.05
		untrained	9	2.44		
	Item 6	trained	6	2.17	-.167	n.s.
		untrained	9	2.33		

6 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 interdisciplinary teams has been found. Yet, D-School teams consistently report less difficulty 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. E. g., 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 interdisciplinary 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 work-shop. 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.

7 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 mod- els? 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 interdisciplinarity 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 interdisciplinarity 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 interdisciplinary instead of monodisciplinary teams. Or, to address another likely reasoning: Multidisciplinarity may have a considerable positive effect, but not in all contexts. E. g., it comes to unfold its positive impact only after longer periods of time (months, not days). Another viable thought may be that interdisciplinary de- sign 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 wouldn't be design thinkers if we were to duck out of the test, would we?

Acknowledgement

The authors are grateful to Julia von Thienen for her excellent work in the e.valuate team of the Hasso Plattner Design Thinking Research Program.

References

Brown, T.: Design Thinking, in: Harvard Business Review, June 2008, p. 84 – 92.

Change Management Education: An Interview and Discussion, in: Academy of Management Learning & Education, 2006, 5(4), p. 512-523

Cross, N.: The nature and nurture of design ability, in: Design Studies, 1990, Vol. 11, No. 3, p. 127-140, citation p. 130.

Dunn, D.; Martin, R.: Design Thinking and How It Will

Lawson, B.: How Designers Think (4th ed.). Oxford 2006.

Lindberg T., Noweski, C., Meinel, C.: Evolving Discourses on Design Thinking, Technoetic Arts Journal, Volume 8, Issue 1, 2010

Martin, R.: Design of Business, Harvard Business School Press, Mcgraw-Hill Publ.Comp., 2009

Plattner, H.; Meinel, Christoph; Weinberg, Ulrich: Design Thinking. Munich 2009

Rowe, P. G.: Design Thinking, 1991, MIT Press

Von Thienen, J. et al.: The Co-evolution of Theory and Practice in Design Thinking – or: “Mind the oddness trap!”, in Meinel, C. and Leifer, L. (editors): Design Thinking: Understand - Improve - Apply (Understanding Innovation), Springer, Berlin, to appear 10/2010