Dimensions of Creative Evaluation: Distinct Design and Reasoning Strategies for Aesthetic, Functional and Originality Judgments

Bo T. Christensen

Copenhagen Business School, Copenhagen, Denmark
bc.marktg@cbs.dk

Linden J. Ball

University of Central Lancashire, Preston, UK
LBall@uclan.ac.uk

Abstract*: The datasets provided as part of DTRS-10 all relate to what may broadly be labeled as 'design critiques' in an educational context. As such, we chose to center our theoretical analysis on the evaluative reasoning taking place during expert appraisals of the design concepts that were being produced by industrial design students throughout the design process. This overall framing for our research allowed us to pursue a series of research questions concerning the dimensions of creative evaluation in design and their consequences for reasoning strategies and suggestions for moving further in the creative progress. Our transcript coding and analysis focused on three key dimensions of creativity, that is, originality, functionality and aesthetics. Each dimension was associated with a particular underpinning 'logic' that determined the distinctive ways in which these dimensions were seen to be evaluated in practice. In particular, our analysis clarified the way in which design dimensions triggered very different reasoning strategies such as running mental simulations, or making suggestions for design improvement, ranging from definitive 'go/kill' decisions right through to loose recommendations to continue to work on a concept for a period of time without any further directional steer beyond this general appraisal. Overall, we believe that our findings not only advance a theoretical understanding of evaluation behaviour that arises in design critiques, but also have important practical implications in terms of alerting expert design evaluators to the nature and consequences of their critical appraisals.*

Keywords: Design critique, design reasoning, design evaluation, evaluative practices, mental simulation, design judgment

1. Introduction

Evaluative practices are an important aspect of all creative industries, where key individuals are invited to comment on and evaluate products 'in-the-making' during initial creative stages as well as products that are finalized and ready to be communicated to the market (Moeran & Christensen, 2014). Most creative industries have therefore formalized specific roles both for individuals who are helping to advance the creative process and for domain experts who are evaluating the final outcome at gates, reviews or screenings. In this respect the 'design critique' that is a key feature of design education can be viewed as a friendly critical appraisal based

around interactions between designers aimed partly at evaluating the potential, novelty and value of the product in-the-making, but equally importantly as a means to spur the pursuit of new directions, angles and lines of creative inquiry. The critique presents students with opportunities to develop their own design values and preferences and to become more aware of their own design sensibilities (McDonnell, 2014). Design critiques may play out in many different designer relationships, from master-apprentice to peer critiques, using a variety of modalities, including speech, gesture and sketching (see Oh, Ishizaki, Gross, & Do, 2013, for an overview). The outcome of such a design critique may occasionally be a discarded project, but more frequently the critique will initiate a series of further investigations and creative processes in order to strengthen the project.

Within a design critique the dialogue that arises (typically between an experienced designer and one or more less experienced designers) may take the form of an exploratory process that has as its input so-called 'preinventive' structures (e.g., sketches, more or less formalized ideas or concepts, and prototypes), in line with the conceptualization of the creative process offered in the 'Geneplore' model of creativity (e.g., Finke, Ward, & Smith, 1992; see also Finke, 1990). In this respect it is noteworthy that the Geneplore model considers exploratory processes (e.g., contextual shifting and form-before-function reasoning) as being inherently 'creative' in nature. The implication of this view is that exploratory processes should not be overlooked and that the commonly held belief that creativity primarily concerns generation as opposed to exploration is mistaken by virtue of being an overly-narrow conceptualisation of creative activity. We also note that the design critique typically involves a dedicated and formalized role for the design evaluator, who is presented with a preinventive structure to evaluate and to help advance. A typical design critique, then, allows for a relatively clear distribution of roles: (1) a designer (or sometimes a group or team) who has constructed an initial preinventive structure; and (2) another designer (frequently more experienced) who is exploring and evaluating that preinventive structure. The present research utilizes this relatively clear distribution of roles in order to examine the different dimensions of creative evaluation in industrial design education as well as the design strategies employed to attain elevated levels of creativity. While the present dataset revolved around experienced designers critiquing students, the present analysis first and foremost examined how distinct evaluation type logics affect the reasoning and progression suggestions of the experienced designer.

In relation to the theme of creativity we note that the literature has tended to reach a consensus that for a product to be deemed to be 'creative' it needs to display the properties of both novelty and usefulness to some domain (e.g., Amabile, 1996; Mayer, 1999). While novelty is typically seen as the hallmark of creativity, the arguments for including a second dimension revolve around the observation that originality is not enough: schizophrenic ramblings, although novel, are not in themselves creative as they lack domain value or usefulness. The creative property of 'usefulness' or 'domain value' is, however, conceptually vague, and needs further specification in order to make sense in any concrete domain. For the design domain, Nelson and Stolterman (2012) have listed multiple important judgment types operating under what they term 'design judgments'. While they do not claim to have derived an exhaustive list, sample types include framing judgments, appearance judgments, quality judgments, compositional judgments, and navigational judgments. For the present purposes, we wish merely to illustrate that evaluative types differ in terms of the underlying evaluation logic, leading to differences in reasoning strategies and proposed ways forward in the design process. So, for the present purposes it

suffices to claim minimally that two high level and important values in industrial design are functional value and aesthetic value. Note, however, that there may well be other high level evaluation dimensions in industrial design than the ones we have chosen here, and that the chosen dimensions may be separated into more fine-grained sub-categories.

Below we seek to theorize on the nature of these three dimensions of creativity in industrial design (i.e., originality, functionality and aesthetics), and how they may predict differential behavior for the designers and evaluators who are traversing through their creative processes. The previous creativity literature has tended to ignore the question of the 'logic' behind these distinct dimensions of creativity and how this logic may relate to the way in which these dimensions are evaluated in practice. In the context of design, for example, it is clearly possible to evaluate design objects from the perspective of different value systems, such as functional value or aesthetic value, such that the actual 'process' of reasoning about value may take several distinct forms. In the present research we sought to explore such different types of reasoning and the progression of design ideation (if any) that takes place in evaluating 'functional value' (e.g., usability), 'aesthetic value' (e.g., visual form), and 'originality value' (e.g., consumer-perceived novelty or domain changing potential). While these value dimensions may frequently be entangled in practical creative evaluation (with multiple dimensions co-occuring, some foregrounded and others backgrounded in concrete evaluative statements), in the present study they are analyzed as distinct entities in their pure form in order to draw out their core differences and respective relation to reasoning strategies.

1.1 The logics of creative dimensions

Originality evaluation

Evaluations of originality assume that ideas and products exist in objective temporal reality, and that it is possible to analyze the history and development of concepts. Value is placed especially on domain-specific originality that may later spark off multiple, fruitful variants in the domain in question in a germ-like manner. This implies a heavy emphasis on the value of a design arising from its being the 'first' of a (new) kind. Given that originality is basically seeking novelty of kind, dismissal of a design due to lack of originality should frequently lead to a rapid rejection of the whole concept, rather than leading to suggestions on how to improve the concept's originality. In other words, an unoriginal concept needs to be discarded, rather than developed. Two modes of originality judgments may exist, one valuing the perceived originality by consumers (e.g., as practiced by marketers; see Dahl & Moreau, 2002; Moldovan, Goldenberg, & Chattopadhyay, 2011), while the other values the factual originality for the domain in question (e.g., as practiced by domain gatekeepers or experts; cf. Czikszentmihalyi 1990; Amabile, 1982). This logic of the dimension of originality ties it closely to 'go/kill' design decisions for whole concepts. In a design process such evaluations and decisions revolve around the birth of ideas, and are made in the early stages of the design process.

Functional evaluation

Functional evaluation assumes an objective physical reality against which a design concept may ultimately be tested. Much functional evaluation involves mentally 'simulating' whether the prescribed requirements are met to a satisfactory degree, and whether the design object performs

as specified. A mental model 'run' is a change made to a mentally constructed 'model' that allows for reasoning about new possible states (e.g., see Ball & Christensen, 2009; Ball, Onarheim, & Christensen, 2010; Christensen & Schunn, 2009; Wiltschnig, Christensen, & Ball, 2013). As a consequence, a great deal of functional evaluation focuses on detecting and resolving errors or shortcomings of design elements. While much evaluative design dialogue may revolve around mentally reducing functional uncertainty and turning that uncertainty into approximate answers (e.g., see Ball & Christensen, 2009; Christensen & Schunn, 2009), ultimately the real challenge for functional value is whether the design object operates as required when put to the test in the laboratory or in real-world trials and experiments. As such, functional design evaluation is fundamentally distinct from socially-oriented consensual agreement that is described in much of the creativity evaluation literature (e.g., Amabile, 1996; Czikszentmihalyi 1990) given the insistence that physical reality remains the ultimate challenge for the functional value of design ideas. Functional evaluation will frequently lead to identification of mis-behaving sub-parts that may be improved upon in an incremental manner through the design development process. The focus therefore rests on the life of ideas and concepts, that is, design as a process of continual improvement rather than design as a number of units that are simply screened and selected or discarded.

Aesthetic evaluation

While it has been claimed that beauty is fundamentally in the eye of the beholder, research has identified multiple dimensions influencing aesthetic judgments, some relating more clearly to the object in question (e.g., symmetry, complexity and contrast), some to the prevalence of similar objects (e.g., prototypicality and familiarity), some to the classification of the object (e.g., style or content), and some to qualities of the perceiver (e.g., cognitive mastery, expertise, personal taste and interests) with both cognitive and affective dimensions (Leder et al., 2004). Controversies among art appreciation theorists date back millennia, rendering it unwise to make solid claims about the fundamental nature of aesthetics. Nonetheless, certain qualities of aesthetic judgments in industrial design may be highlighted when making comparisons to functionality and originality judgments. In particular, aesthetic evaluation seems to have a much clearer emotional or hedonic tone compared to judgments of originality or functionality.

Given that important dimensions of aesthetic evaluation rest on qualities of a particular perceiver (an individual) or a particular class of perceivers (a social or cultural group), then the possibility for variance in taste can presumably be considered higher for aesthetic evaluation compared to the other two types of evaluation used here. Likewise, aesthetic evaluation may be subject to greater temporal shifts in appreciation (i.e., in line with the existing social consensus relating to taste or style). Finally, compared to the other evaluation types, aesthetic evaluation rests to a larger degree on the affective and cognitive dimensions associated with perceiving the object. The actual perceptual performance seems less important in evaluating originality and functionality, whereas one has to perceive the object with one's own senses in order to judge aesthetic pleasure. This also implies that judging the aesthetic pleasure of non-perceptual ideas (e.g., designed objects only conveyed through words) is extremely difficult. Materiality matters to aesthetic appreciation, both to the evaluator of aesthetic objects, but equally so to the creator in the creative process, where the actual construction and interaction with the designed object is important as the object 'talks back' as it takes shape.

1.2 Propositions and hypotheses

Given the qualities of the three evaluative types that have been selected for the present analysis it is important to question what design strategies might be applied in relation to each of these evaluation types. What might we expect in terms of reasoning and suggestions for design idea progression for each of these evaluation types? Based on the aforementioned descriptions of the core differences between the evaluation of originality, aesthetics and functionality, three basic propositions were derived that contextualized the present analysis, as follows:

1. The three types of evaluation diverge on what may be described at the ontological basis of the evaluation. Here functionality evaluation stands out given the ability ultimately to test and simulate the capacity for the design to meet certain objective criteria or requirements. Admittedly, functionality evaluation may sometimes be assessed against more subjective criteria, such as the usefulness of the design, but the important point here is that frequently function is a matter of objectively testable threshold values. As such, functionality evaluation should more frequently lead to suggestions for experimentation and testing of the design when compared to either originality evaluation or aesthetic evaluation. Furthermore, as a mental shortcut to replace detailed experimental testing it would be expected that mental simulation of proposed designs would be used as a heuristic strategy.

2. The three evaluation types diverge when it comes to the conception of what an 'idea' entails in creative or innovative processes. In general, creativity theories dissociate in terms of whether ideas are perceived as 'units' or as 'processes'. The inclination in originality judgments is for designers to identify and compare designs as 'entities' whilst looking for novel concepts, which may be contrasted with the procedural understanding of design that is particularly sought in functionality evaluation, but also in aesthetic evaluation, where designs are viewed mainly in terms of continuous development. While originality evaluation maintains a focus on the birth of ideas and the choice amongst alternative design entities, we contend that aesthetic and functionality evaluation focus on the life of ideas, and the continual improvement of design though the development of elements by means of additions and changes.

3. Finally, the three evaluation types diverge in terms of the importance of the perception of the design object as well as interaction with the object during the evaluation process. Aesthetic evaluation stands out in this respect in that aesthetic evaluation or aesthetic development seem to demand direct perceptual interaction with the design object in question, especially in order to be able to draw out the emotional responses to the object. This 'need' may spill over into strategic suggestions for advancing design improvements in that further recommendations may be given to continue design development even without specific guidance as to which particular parameters to change. That is, a concept is perhaps more likely to be identified as having 'potential' or to be 'of interest' in relation to aesthetic judgments, without the ability to verbalize exactly how, or in what direction, the concept should be taken. Similarly, it may be more difficult in aesthetic evaluation than in functional evaluation to mentally simulate variations of a design, particularly

in light of the difficulty to pick up on the hedonics or emotional tone of a design merely on the basis of non-physical and non-sketched ideation.

These three aforementioned propositions as to how originality, functional and aesthetic evaluation differ can be rephrased as specific hypotheses for each possible evaluation pairing, as follows:

- Comparing aesthetic evaluation to functionality evaluation we predict in the former more suggestions for development through trial and error (H1a), less mental simulation (H1b) and fewer suggestions for testing the concept (H1c).

- Comparing originality evaluation to aesthetic evaluation we predict in the former less mental simulation (H2a), more 'go/kill' decisions for whole concepts (H2b) and fewer suggestions for development through trial and error (H2c).

- Comparing functionality evaluation to originality evaluation we predict in the former more suggestions for changing elements or forms (H3a), more mental simulation (H3b), fewer 'go/kill' decisions (H3c) and more concept testing suggestions (H3d).

In addition to these formal hypotheses, we also wanted to explore potential differences between the three chosen evaluation types in terms of their overall level of epistemic uncertainty, and valence. We believe it is the first time that the logics behind these three types of design evaluation have been theorized upon and compared in design critiques. A further implication of the present argument is that distinct creative domains are likely to diverge in the proportions of the three chosen evaluation types in actual design practice. As argued by other papers from the DTRS10 symposium, the literature on how creativity and creative evaluation varies across disciplines is sparse (Mann & Araci, 2014; Yilmaz & Daly, 2014). Examining the differential proportions of originality evaluation, aesthetic evaluation and functional evaluation across domains is, however, beyond the scope of the present paper, but we nevertheless believe it highly likely that these types of logics may help explain differences in creative evaluation practice, for example, between artistic domains and technical or scientific domains.

2 Methods

The present study focused on the coding and analysis of design critique data from undergraduate and graduate industrial design courses at a public university deriving from the DTRS-10 dataset (Adams & Siddiqui, 2013). The data that we analysed consisted of 13 supervisor/student interactions across 39 transcripts, covering all stages of the design process within an educational setting (i.e., first review/D-search; second review/concept review; client review; look like/concept reduction; final review). The data were segmented according to turn-taking during spoken dialogue, resulting in a total of 4316 segments, ranging from 108-717 for each student, and 19-470 for each design critique session. Below we describe the detailed approach that we adopted to code the transcripts.

2.1 Transcript coding

The transcribed industrial design critiques were independently coded by three student coders who were unaware of the hypotheses underpinning this study. Each student coder coded a subset of the data. The coders were first trained in the analysis of verbal transcripts and were familiarized with the videos and datasets. They then applied six different codes during five iterations of going through the datasets.

2.2 Coding of evaluation episodes

Initially, all statements were identified involving evaluations that were uttered by the evaluator (i.e., in the present dataset, a senior designer). For the purposes of this analysis a statement of evaluation was defined as any statement that comments on or that evaluates (either positively or negatively) the designed product or a design idea. The coding excluded any evaluations that commented on the design process, on presentation techniques (e.g., PowerPoint visuals of no importance to actual design ideas) or on the capabilities of the student designer (so long as these were also unrelated to the designed object). In this way the focus of the coding was specifically on the evaluation of design products or ideas. Examples of comments that were coded as statements of evaluation included: 'that's cool', 'great idea', 'I don't like the x component' and 'this bit might not work'.

Following the identification of a 'statement of evaluation', a block of segments relating to this evaluation was identified, which contained descriptions and/or explanations of the design idea (usually uttered before the statement of evaluation) as well as segments involving further development or reasoning concerning the evaluation (usually uttered after to the statement of evaluation). An episode of evaluation was then coded, covering both the design explanation, the statement of evaluation, and the reasoning/development taking place subsequently. In principle, a single segment could be coded as an episode in itself, but most typically an episode spanned multiple segments.

Coding of evaluation valence

All statements of evaluation (see above) were coded in a binary manner for their valence, that is, they were designated as possessing a positive or a negative valence (see Tables 1 and 2 for examples). In situations where the statement of evaluation contained both positive and negatively valenced utterances, then the evaluation episode was coded as 'both positive and negative'.

Table 1. Transcript extracts that show positively valenced evaluations

Speaker	Discussion
	(Undergraduate; Addison; Final review; line 26)
Simon:	But I kinda like – I don't know what to call it underwear or bikini or whatever you wanna call that –
	(Undergraduate; Lynn; First review; line 19)
Gary:	Excellent, excellent.

	(Undergraduate; Todd; First review; Line 28)
Gary:	Yeah, this is, this is pretty neat. This would be great. This would probably be fiberglass or molded plastic.

Table 2. Transcript extracts that show negatively valenced evaluations

Speaker	Discussion
	(Undergraduate; Lynn; First review; line 121)
Gary:	Um, the bad thing about these is these, these actually – um, may not really be too stable, though, you know –
	(Graduate; Eva; Concept review; line 22)
Simon:	I missed the anti-gravity. Where is it? Oh, vacuum environment. But a vacuum environment doesn't make things float.
	(Undergraduate; Lynn; Client review; line 9)
Darren:	Wha-, well, personally, personally I don't see that once again, I don't see that as a marketable model. I don't think it will be used in the way you think it is.

Coding of evaluation types

All statements of evaluation were also coded for whether they pertained to design aesthetics, to design function or usage or to the originality of the design. Evaluations relating to design appearance or form were coded as aesthetic evaluations (e.g., as arising in relation to the look, feel or smell of the designed object; see Table 3 for examples). Evaluations relating to design usage or technical function were coded as functionality evaluations (e.g., 'this functional element needs to be changed', 'it's probably not going to work' or 'users will probably not appreciate this element'; see Table 4 for examples). Evaluations relating to the distinctiveness or novelty of the design were coded as originality evaluations (e.g., 'this has been seen before', 'this design is unique', 'it's radically different', 'this is the safe option' or 'the design is quite different'; see Table 5 for examples).

Table 3. Transcript extracts that show examples of aesthetic evaluations

Speaker	Discussion
	(Undergraduate; Addison; Client Review; line 16)
Darren:	Well you've got a very different, uh, progression from what we see in the top to the bottom. I think they're both valid. You know I guess my question was - was that part of your thought

	process because both forms are really nice?
	(Undergraduate; Alice; 2nd review; line 121)
Gary:	This was – save this for another – this one's kinda neat. I really loved how this curved around.

Table 4. Transcript extracts that show examples of functionality evaluations

Speaker	Discussion
	(Graduate; Mylie; Client review; line 92)
Peter:	Ya' know, I love the idea of having accessories that, that can hang from the branches that allow you to customize it and, ya' know, it supports different functionality.
	(Graduate; Walter; Concept review; line 362-363)
Simon:	Yeah, the water will be everywhere and there's no point in - why even have it then? But I do like it as three separate containers and -
	... your, your basket is what goes into the machine. it could be - course you always have to wash all - you have to wash everything.

Table 5. Transcript extracts that show examples of originality evaluations

Speaker	Discussion
	(Undergraduate; Alice; 2nd review; line 66)
Gary:	– medium, and extreme to some degree. That's, that's kinda it helps them. So this is if you wanted to design something really similar to what everybody else has done, this is what I'd recommend. But your goal as a designer is – they're not hiring you to, to, ah, to analyze the market. I mean they're doing that. They're not analyzing – they're not hiring you to do CAD.
	(Graduate; Eva; Client review; line 77)
Chuck:	This one seems a little far-fetched. I mean, like I – like I said, I appreciate the, uh, I appreciate the out, ya' know, the thinking outside the box, but it's, I mean, maybe we're too – in too much reality.

Coding of mental simulation

The codes pertaining to the presence of mental simulation were based on those developed by Christensen and Schunn (2009; see also Ball & Christensen, 2009; Ball et al., 2010; Wiltschnig et al., 2013), which were, themselves, adapted from research reported by Trickett and Trafton (2002; see also Trickett and Trafton, 2007). Within this coding scheme a mental model 'run' is viewed as being a mentally constructed 'model' of a situation, object or system of objects that is

grounded either in the designer's memory or in the designer's mental modification of design objects that are physically present. As such, mental simulation enables designers to reason about new possible states of a design object in terms of its qualities, functions, features or attributes, but without the need for actual physical manipulation of the object itself. It should be noted that mental simulations are not merely limited to technical design properties, but can also relate to imagining other kinds of dynamic situations relating to the designed object. Such situations might extend to envisaging changes arising from end-user interactions with the object or to imagining an individual's aesthetic appreciation in relation to altered aspects of the object.

Whatever its end goal, the key feature of a mental simulation is that it involves a simulation 'run' that alters a mental representation to produce a change of state (e.g., Trickett and Trafton, 2007; see also Richardson & Ball, 2009). What this means is that a mental simulation necessitates a specific sequence of representational changes, commencing with the creation of an initial representation, progressing to the running of that representation (where it is transformed by additions, deletions and modification), and finishing off with a final, changed representation (e.g., Christensen & Schunn, 2009). These three components of the mental simulation (i.e., the initial representation, the simulation run, and the changed representation) are not conceptualised as being mutually exclusive, but can occur in the same transcript segment, although typically they extend over several segments. Examples of mental simulations are shown in Table 6.

Table 6. Transcript extracts that show examples of mental simulations

Speaker	**Discussion**
	(Undergraduate; Adam; 2nd review; line 35-37)
Gary:	Yeah, and then you've got this sort of element. Now one of things when it goes on the floor, um, you may consider maybe that's a have some semi-soft machinable plastic pieces of material. Um, or maybe it could be, um, a – maybe a metal piece or something. I don't know. But, anyway, we need to have some kind of structure. You won't, you won't have narrow enough fabric to the floor – even if slightly, maybe like wood. Um, so then this, this could be uh wood piece that could be, could be fabric in here maybe it comes down, or something just, keep just, just keeps the [clears throat] fabric from touching the floor and it's already kind of moisture or whatever at least it's, maybe it could be waterproof or more durable. Otherwise, you – again, and this could, this could just be like three-quarter, half inch, but something you never see because maybe step it back a little bit and be – maybe something that – and these details you can work out later.
	(Undergraduate; Sheryl; Look like; line 68-72)
Gary:	Well, I'd get some stretch fabric to where you maybe hide 'em back on the side – on the inside.
Sheryl:	Oh, yeah, like – oh, what is that fabric called that you see these like book covers with? Do you know what I'm talking about – high school? [Laughs]
Gary:	No, but I mean go to a fabric store and get the stretchiest fabric you can get, ah –

Sheryl:	Okay.
Gary:	And, and you realize that maybe it's time – maybe it's the bottom where you pull everything in and then you, you, ah, you hot melt glue it or something. In fact, you may want to – you build that where, you know, the bottom piece of your multiple layers of cardboard up a little higher, so that way – 'cause fabric's always, always gonna gather. So maybe you, you have a little bit of play in there, maybe a half-inch on the bottom that you could bring it under and say, well, this is, this is just for decorative. Obviously, they will figure out how to make it work.

Coding of epistemic uncertainty

Epistemic uncertainty refers to a metacognitive state that arises during a design process on occasions when a designer is unsure about some aspect of their on-going design work such as their understanding of elements of the problem or their confidence in the effectiveness of solution ideas (e.g., see Ball & Christensen, 2009). Previous design research has demonstrated that the manifest expression of epistemic uncertainty by designers is often associated with strategic shifts in behavior such as increases in mental simulation and analogising (e.g., Ahmed & Christensen, 2009; Ball & Christensen, 2009; Christensen & Schunn, 2007, 2009) as well as increases in problem–solution co-evolution activity (Wiltschnig et al., 2013).

In the present analysis the coding of epistemic uncertainty was achieved using a syntactic approach adapted from Trickett et al. (2005) and Christensen and Schunn (2009) which makes use of 'hedge words' to search for segments within the transcript that contain expressions of uncertainty. In the present analysis these hedge words included terms like 'probably', 'sort of', 'guess', 'maybe', 'possibly', 'don't know', and 'believe'. Text segments containing these words or phrases were located and were coded as 'uncertainty present' if it was also apparent that the hedge words were not being used by the speaker merely as politeness markers (see Table 7 below, which shows extracts from the transcripts where uncertainty was present). Any segment that were not coded as 'uncertainty present' was coded as 'uncertainty absent'

Table 7. Extracts from the transcripts where uncertainty was present (as designated using bold and underlined font)

Speaker	**Discussion**
	(Undergraduate; Alice; 2nd review; line 85)
Alice:	Okay. 'Cause here, I was playing with this idea of having [unintelligible] **think**, and then **maybe** it could be, could be upside down.
	(Undergraduate; Esther; Look like; line 106)
Gary:	So you **probably** want to do that, 'cause you can build up your layers and then you'll need something else.
	(Undergraduate; Lynn; First review; line 172)
Gary:	Okay. And **maybe** there's, **maybe** there's some simple geometry. You gotta – **maybe,**

	maybe it's more straight –
	(Undergraduate; Sheryl; Look like; line 83)
Sheryl:	**I don't know.** That's what I was gonna ask. What do you think is best?

Coding of design idea progression suggestions

This set of codes captures suggestions for progression of design ideas that are made when an experienced designer evaluates one or more design concepts. Each segment of the transcript was assessed in terms of whether it contained a design idea progression suggestion (DIPS) by the experienced designer. Five distinct types of DIPS were coded, as follows:

- **Go/kill idea:** This arose whenever one or more ideas were selected or highlighted as having more or less potential over other ideas (e.g., 'go with this idea'; 'go with these two ideas, but not this one', 'kill idea 3'; see Table 8).

Table 8. Transcript extracts that show examples of 'go/kill' DIPS

Speaker	Discussion
	(Graduate; Julian; Client review; line 29)
Peter:	I think you have other stronger concepts.
	(Graduate; Sydney; Client review; line 21)
Peter:	Okay. Uh, I would say you're probably gonna do 41. Can you go back to that slide?
	(Graduate; Walter; Client review; line 99)
Peter	Those are the two I think strongest ones.

- **Change element or form:** This occurred when a functional or form element was added, removed, or changed for a particular concept or idea (e.g., 'please change the base to another kind of material', 'I would drop this particular bit of your idea', 'you should consider adding this bit', 'these dimensions should be scaled', 'why not add some color to this bit'; see Table 9).

Table 9. Transcript extracts that show examples of 'change form or function' DIPS

Speaker	Discussion
	(Graduate; Mylie; Client review; line 60)
Chuck:	And, ya' know, maybe you add the fragrance thing in and kinda' take it from there.

	(Graduate; Sydney; Client review; line 40)
Chuck:	-- you have shown on the left. It'll probably be a smaller type thing and the air can come from the dryer when you're, ya' know, when you're drying the other clothes. That could be cool and it could just, ya' know, the hot air could kinda' come up and help, help dry those clothes.
	(Graduate; Julian; Client review; line 61)
Peter:	It could be something smaller.

- **Test concept:** This arose when the experienced designer suggested testing the concept (e.g., through experimentation or by testing it on users; see Table 10).

Table 10. Transcript extracts that show examples of 'test concept' DIPS

Speaker	**Discussion**
	(Undergraduate; Todd; Look like; line 65)
Gary:	Talking about get a dowel and drill through the – drill through the bottom all the way up, and, and then, ah, with a drill press and then, ah, gotta dowel and see if it actually functions.
	(Graduate; Julian; Client review; line 52)
Peter:	So I, I would do, ya' know, I, I would concentrate on this, but I, I don't think it's as easy as what you have drawn here with the variation in clothing, it's gonna take some, ya' know, it's gonna take some experimenting on your side.

- **Search for more information:** This was when the experienced designer suggested searching for new or additional information for the design (Table 11).

Table 11. Transcript extracts that show examples of 'search for more information' DIPS

Speaker	**Discussion**
	(Graduate; Julian; Concept reduction; line 157)
Simon:	Okay. So you gotta do a little research.
	(Graduate; Sydney; Client review; line 28)
Peter:	Okay? That's a – that's a – I mean, that's something different that at least I haven't seen. Again, you might wanna look out there. Just Google search or patent search foldable hangers you might see there. I think there's a lot of people that could benefit from something like this and it seems so simple and elegant a solution.

- **Trial and error**: Thus occurred whenever the experienced designer asked the student to play with the concept, try out different things, or work on the concept for a specified time, without further specifying what the outcome might be (e.g., 'play with it', 'play

with the dimensions a bit', 'try different things out', 'work with it for a few hours'; see Table 12).

Table 12. Transcript extracts that show examples of 'trial and error' DIPS

Speaker	Discussion
	(Undergraduate; Alice; 2nd review; line 177)
Gary:	So play with your forms and dimensions, and then these others which are really, really exciting as independent pieces, that's really refreshing. Both these are really fun. Both of 'em have great merit. [Clears throat] This, um, you could play around with the height on this thing.
	(Undergraduate; Lynn; First review; line 184)
Gary:	But, again, you, you've got – you've – I'll give you my input and you're the designer. If you're passionate about something and, ah, you could appropriate the time for it, just go for it. This is something that you really like, so take it to a level, but I would maybe spend a couple of hours on it, trying to dial in the geometry.

Inter-coder reliability checks

In order to undertake a reliability check of the transcript coding we selected a set of transcripts of interactions between a single student and supervisor, which covered three sessions (client review, look like and final review). The transcripts involved a total of 210 segments (i.e., approximately 5% of the full dataset). Two individuals coded the transcript independently, and reliability was then estimated using Cohen's Kappa measure. In case of insufficient reliability, the coding scheme was revised, the coders re-trained, the data re-coded, and a new round of reliability checking was conducted. Following the achievement of sufficient reliability, all disagreements were resolved through discussion between the coders. As shown in Table 13, all codes reached a satisfactory level of inter-rater agreement. Mental simulation and design idea progression suggestion can be characterized as 'fair-to-good' agreement, while the remaining codes had excellent inter-coder agreement according to the rule-of thumb provided by Fleiss et al. (1981; see also Fleiss, 1981).

Table 13. Kappa coefficients for inter-coder reliability

Code	Kappa coefficient
Mental Simulation	.71
Evaluation Episodes	.75
Design Idea Progression Suggestion	.68
Evaluation Valence	.86
Evaluation Type	.85
Uncertainty	.90

3 Results

3.1 Evaluation episodes

Across the transcripts we identified 157 unique evaluation episodes, which ranged from 1 to 49 segments, averaging 9.9 segments per episode. Evaluation episodes thus made up 36.2 % of the segments in the transcripts, which is not surprising given that the essence of design critique is centrally focused on the evaluation of concepts. Following each student across the sessions in the design process showed that evaluation episodes received by each of the students ranged from 0 to 32, with an average of 12.1 episodes per student.

3.2 Evaluation types

Of the 157 evaluation episodes, 42% pertained to aesthetic evaluation, 46.5% to functional evaluation, and 11.5% to evaluation of the originality of concepts. A chi-square analysis of the distribution of the three types of evaluation by session (Figure 1) was prohibited due to the presence of expected counts less than 5. As a consequence, the final review (Session 5) was excluded from the analysis and Session 3 (client review) and Session 4 (look like; concept reduction) were merged into a single session (see Table 14). The resulting chi-square analysis revealed significant differences in the distribution of evaluation types by session, χ^2 (4) = 18.34, $p < .001$. Follow-up 2 x 2 chi-square tests revealed that when comparing the first session to later sessions, originality evaluations, χ^2 (1) = 7.47, $p < .007$, and aesthetic evaluations, χ^2 (1) = 12.45, $p < .001$, arose more frequently in the first session than later sessions relative to functionality evaluations. However, aesthetic evaluations and originality evaluations did not differ from one another in this respect, χ^2 (1) = 0.01, ns.

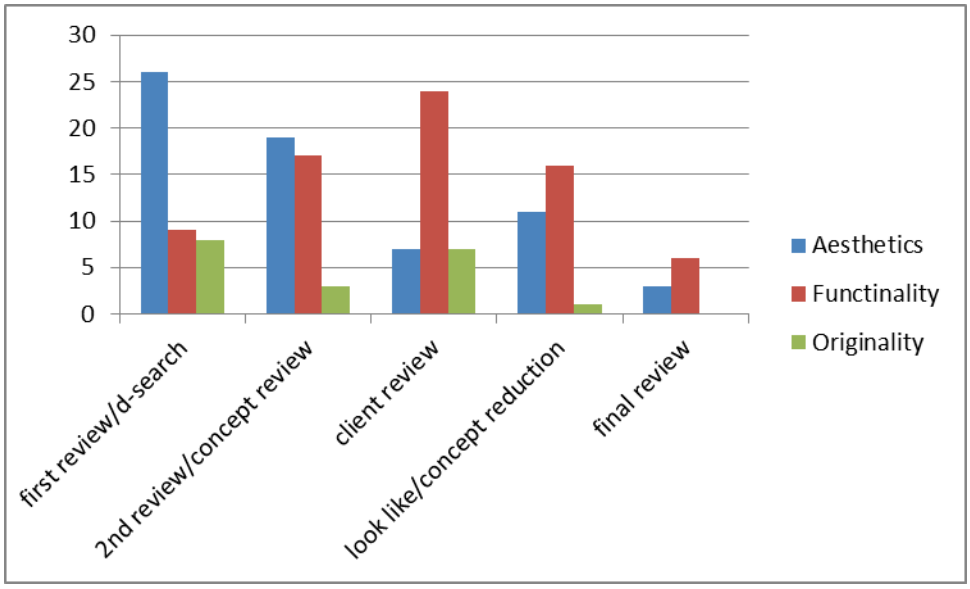

Figure 1: The frequency of evaluation types across sessions

Table 14. Contingency table showing the frequency of evaluation types by session (note that Session 5 was omitted from the analysis while Sessions 3 and 4 were combined in order to apply a chi-square test)

	Session		
Evaluation Type	**1**	**2**	**3 + 4**
Aesthetic	26	19	18
Functionality	9	17	40
Originality	8	3	8

3.3 Evaluation valence

Of the 157 evaluation episodes, 69.4% were positively valenced, 15.9% were negatively valenced, and the remaining 14.7% of the episodes contained both positive and negative evaluations within the same episode (see Table 15 for frequency data). When excluding episodes containing both positive and negative evaluations it was observed that evaluation types differed significantly in terms of their valence, χ^2 (2) = 24.76, $p < .001$. Subsequent 2 x 2 Fisher's exact tests revealed that aesthetic evaluations ($p < .001$) and originality evaluations ($p < .004$) were significantly more often positive (indeed, almost entirely so) when compared to functional evaluations, while aesthetic evaluations and originality evaluations did not differ from each other in their valence ($p = 1.00$). The surprisingly large proportion of positively valenced evaluative statements (given the context of a design critique session) may be seen in the light of Oak and Lloyd's (2014) key point that during a critique the institutional context, the associated roles of participants, and the management of face, all contribute to shaping what can be said and how it is said. As Oak and Lloyd show in detailed analyses of single critique encounters, the instructor Simon maintains a rather explicit vocabulary of what is to take place during the design critique (stating that he 'tore into the students work'), which is somewhat in contrast to the somewhat gentle remarks actually offered during the critique.

Table 15. Contingency table showing the frequency of evaluation types by evaluation valence

	Valence		
Evaluation Type	**Positive**	**Negative**	**Total**
Aesthetic	56	2	58
Functionality	38	21	59
Originality	15	0	15
Total	109	23	132

3.4 Epistemic uncertainty

The transcripts contained a total of 751 segments with epistemic uncertainty present, amounting to 17.4% of the data. For each individual student/evaluator pair, there was an average of 57.8 segments with uncertainty present, ranging from 18-119 uncertainty segments per pair. A one-way ANOVA revealed that the three evaluation episode types did not differ significantly in terms of their level of epistemic uncertainty, $F(2, 156) = .488$, $p = .62$.

3.5 Mental simulation

A total of 113 mental simulations were identified across the transcripts. For each individual student/evaluator pair, an average of 8.9 mental simulations were carried out, ranging from 0-18 mental simulations per pair. Simulation segments occurred much more frequently inside evaluation episodes than outside (Table 16), attesting to the tight coupling between mental simulations and evaluation episodes in the present transcripts, $\chi^2 (1) = 415.29$, $p < .001$). Only 15 of the 113 mental simulations did not relate to an evaluation episode in at least one segment.

Table 16. Contingency table showing the number of segments when simulation was resent and when simulation was absent within evaluation episodes versus outside evaluation episodes

	Within evaluation episode	Outside evaluation episode	Total
Simulation present	343	78	421
Simulation absent	1217	2678	3895
Total	1560	2756	4316

As has been found previously (e.g., Ball & Christensen, 2009; Christensen & Schunn, 2009; Ball et al., 2010; Wiltschnig et al., 2013), the analysis of the present transcripts revealed that mental simulations were run in situations of elevated epistemic uncertainty. Simulation segments thus contained epistemic uncertainty far more frequently than non-simulation segments, $\chi^2 (1) = 105.07$, $p < .001$ (Table 17).

Table 17. Contingency table showing the number of segments when simulation was present and when simulation was absent that revealed the presence versus absence of uncertainty

	Uncertainty present	Uncertainty absent	Total
Simulation present	149	272	421
Simulation absent	602	3293	3895
Total	751	3565	4316

3.6 Design idea progression suggestions

Across the evaluation episodes there were a total of 153 design idea progression suggestions (DIPS) within episodes. These were distributed as follows: 45 go/kill DIPS; 67 changes to form or function DIPS; 10 test concept DIPS; 9 search for information DIPS; and 22 trial and error DIPS. To examine whether the three evaluation types differed in terms of progression suggestions and mental simulation runs we applied logistic regression analyses. Logistic regression enabled us to predict the probability that an evaluation type was linked to a particular type of DIPS or to the occurrence of mental simulation. The predictor variables were therefore the five DIPS (i.e., go/kill; change form or function; test concept; search for information; trial and error) as well as mental simulation, with all predictor variables coded dichotomously. In order to test the hypotheses, three binary logistic regression models were run for each evaluation type pair, as described in the following sub-sections.

Modeling aesthetic to functionality evaluation types

For the aesthetic and functionality evaluation pair we carried out a stepwise regression (Wald forward), which left two variables in the final equation (i.e., test concept DIPS and trial and error DIPS). An evaluation of the final model versus a model with intercept only was statistically significant, χ^2 (2, N = 138) = 13.03, $p < .001$. The model was able to classify correctly with an overall success rate of 58%. Table 18 shows the logistic regression coefficient, Wald test, and odds ratio for each of the final predictors. The odds ratio indicates that a functional evaluation compared to an aesthetic evaluation is 23.55 times more likely to suggest testing the concept and 4.37 (i.e., 1/0.23) times less likely to request trial and error behavior along the lines of playing with the concept.

Table 18. Logistic regression (final model) predicting evaluation type (aesthetic vs. functional) from design idea progression suggestions and mental simulation

		B	SE	Wald	df	Sig	Exp(B)
	DIPS–Test concept	3.16	1.20	6.93	1	.01	23.55
Step 2	DIPS–Trial and error	-1.47	0.67	4.82	1	.03	0.23
	Constant	0.11	0.19	0.35	1	.56	1.12

Modeling aesthetic to originality evaluation types

For the aesthetic and originality evaluation pair we again carried out a stepwise regression (Wald forward), leaving two variables in the final equation (i.e., go/kill DIPS and mental simulation). A test of the final model versus a model with intercept only was statistically significant, χ^2 (2, N = 85) = 10.16, $p < .007$. The model was able correctly to classify with an overall success rate of 78% (see table 19) The odds ratio indicates that an aesthetic evaluation compared to an

originality evaluation is 3.28 times less likely to suggest selecting or killing the concept and 3.70 (i.e., 1/0.27) times more likely to be associated with the performance of mental simulation.

Table 19. Logistic regression (final model) predicting evaluation type (aesthetic vs. originality) from design idea progression suggestions and mental simulation

		B	SE	Wald	df	Sig	Exp(B)
Step 2	DIPS–Go/kill	1.19	0.56	4.59	1	.03	3.28
	Mental simulation	-1.31	0.69	3.60	1	.06	0.27
	Constant	-1.37	0.42	10.83	1	.00	0.25

Modeling originality to functionality evaluation types

For the originality and functionality evaluation pair, Stepwise regression (Wald forward) was once again carried out, leaving three variables in the final equation (i.e., go/kill DIPS, search for information DIPS and change form or function DIPS). A test of the final model versus a model with intercept only was statistically significant, χ^2 (3, N = 85) = 20.78, $p < .001$. The model was able correctly to classify with an overall success rate of 82% (see Table 20). The odds ratio indicates that an originality evaluation compared to a functional evaluation is 5.39 times more likely to suggest go/kill decisions by selecting or killing the concept, 15.18 times more likely to suggest searching for more information, and (1/0.114) = 8.77 times less likely to suggest changing elements of the form or function of the design concept.

Table 20. Logistic regression (final model) predicting evaluation type (originality vs functionality) from design idea progression suggestions and mental simulation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 3	DIPS–Go/kill	1.68	0.60	7.96	1	.01	5.39
	DIPS–Search for information	2.72	1.28	4.52	1	.03	15.18
	DIPS–Change form or function	-2.17	0.82	7.05	1	.01	0.11
	Constant	-1.51	0.43	12.40	1	.00	0.22

Collinearity checks

Given our hypothesis that mental simulation should be related more to functionality evaluation than to originality evaluation, it was surprising that mental simulation did not become a significant predictor in the final model reported in the previous analysis. One possible confound in this analysis is that some of the independent variables may display collinearity, in particular

mental simulation and the DIPS of changing a form or functional element. Theoretically these latter two strategies appear to be related since an important aspect of both revolves around a changed conception of an initial representation. Inspection of the correlation matrix for all predictor variables (see Table 21) confirms that mental simulation and the DIPS of changing a form or functional element were indeed highly correlated ($r\varphi = .700$, $p < .01$), evidencing the potential for collinearity confounds.

Table 21. Correlations between predictor variables

	Change form or function	Test concept	Search for information	Trial and error	Mental simulation
Go/kill	-0.148	.008	-.035	.109	-.112
Change form or function		.250**	.230**	.097	.700**
Test concept			.048	.346**	.201*
Search for information				.295**	.123
Trial and error					.102

* $p < .05$; ** $p < .01$

One way of eliminating this possible collinearity effect from the analysis is to remove one of the correlated predictor values. This was done for each of the above three models. Upon removal of the change form or function DIPS, the model for aesthetic evaluation and functionality evaluation showed no difference, yielding identical predictors and effect sizes to those noted above. Removing mental simulation from the model of aesthetic evaluation and originality evaluation did not enable the change form or function DIPS to enter the model, leaving a single variable as a predictor. Finally and most importantly, removing the change form or function DIPS from the model of functionality evaluation and originality evaluation did make a difference, with mental simulation now entering the model at the third step in the regression, subsequent to the go/kill DIPS and search for information DIPS. A test of the final model versus a model with intercept only was still statistically significant, χ^2 (3, N = 85) = 18.87, $p < .001$. The model was able correctly to classify with an overall success rate of 81%. In the model, mental simulation predicted evaluation type with B = 1.979; Wald χ^2 (1) = 5.87, $p <. 02$. The odds ratio indicated that an originality evaluation compared to a functionality evaluation was 7.25 (i.e., 1/0.138) times less likely to be associated with the running of mental simulations.

These additional analyses suggests that specifically (and only) for the model of originality and functionality evaluation it was likely that collinearity with the change form or function DIPS may have masked the fact that mental simulations are run more frequently for functionality than for originality evaluations. As such, correcting for collinearity by removing one of the predictor variables was successful in providing general support for the hypothesis that in functionality evaluations mental simulations are run more frequently than in originality evaluations.

3.7 Qualitative analysis

In this section we present three extended sections of transcript that we have selected for each type of evaluation, which illustrate some of our major findings.

Aesthetic evaluation: 'I think there's something there – spend two hours on that!'

The transcript extract presented in Table 22 illustrates how aesthetic judgments frequently lead to suggestions for the designer either to spend time on the concept or to play with the concept in a 'trial and error' manner without the provision of any further specifications as to what is supposed to be developed or changed. In other words, the evaluator simply leaves it to the student to develop the concept further over a period of time, which in the illustrative example below is for around a couple of hours.

Table 22. An extract showing a positively valenced aesthetic evaluation associated with suggestions for the student simply to spend further time developing the concept

Speaker	Discussion
	(Undergraduate; Lynn; first review; Line 178-189)
Gary:	I don't wanna influence you on the, the curvilinear thing you got going on and an organic shape, but like once you start laying this out scale-wise, then you might find out that maybe your proportions – some of these may not work for you. But again, what I would do, I would use today – I would, I would give yourself about just in the beginning, say you've got five concepts, try to give yourself at least two hours for development this one in terms of height, whatever, you know, whatever you think is – with the requirements of the design brief. Spend a couple hours on that and exhaust every single possibility, and then stop, and then go look at this one, work for two hours –
Lynn:	Mm-hmm.
Gary:	– ah, two or three hours. Then work on another one. But if you find out that one of these – and that, that's why I want you to do some hard stops on this – [clears throat] the ones that are more complicated, if they end up requiring a lot more effort, then maybe that's telling you as a designer, maybe it's maybe too complex for this, this project. Not saying it's a bad idea, but I just wanna make sure that, that, ah, you know, you've gotta because of what we're doing. We're doing accelerated program on this.
Lynn:	Mm-hmm.
Gary:	And, ah, they – this may be applicable on another project, if you have a pro-, time to develop it up. I'd much rather have you develop something you can get done in a, an appropriate amount of time that you feel real good about it, instead of spending a lotta time on something that's just not, doesn't don't quite feel – 'cause you had to spend so much time on the geometry.
Lynn:	Mm-hmm.
Gary:	But, again, you, you've got – you've – I'll give you my input and you're the designer. If you're passionate about something and, ah, you could appropriate the time for it, just go for it. This is something that you really like, so take it to a level, but I would maybe spend a couple of hours on it, trying to dial in the geometry.
Lynn:	Mm-hmm.

Gary:	See how you feel about it, and then take each of these other concepts. Okay, this, you know, this was a good one. That's a good – this would be a good simple independent one, nice small for you. This is pretty close to being done. I mean -
Lynn:	Oh.
Gary:	-it's – that's pretty nice. Yeah, in terms of function This one's pretty close to being fleshed out.

Functionality evaluation: 'This bit may not work; you should perhaps change it, and run some tests'

The transcript extract shown in Table 23 illustrates a functionality evaluation that is associated with the involvement of a mental simulation run in order for the concept to be tested and thence to progress via a changed representation.

Table 23. An extract showing a functionality evaluation associated with a simulation run to test a concept

Speaker	**Discussion**
	(Undergraduate; Todd; Look like; line 50-51)
Todd:	Oh, um, maybe I'll put some fabric on it.
Gary:	That's always safe. Then with, with something with fabric you can see through. Then the nice thing about this, obviously, you, you got ways to – since this kind of rotates you could bring it in and you spray adhesive of the, ah – I'd confirm with your model, your rotation. Ah, and then you can get, you can get the cheap – I would just get the cheap foams, you know, the cheap pink foam, one of the big sheets. Ah, and then obviously when you're seeing, seeing these parts of it – you stack 'em up. Um, obviously, you need to make sure you have – e-, every two-inch section, you gotta have some sort of regist-, registration hole so you can put a dowel in when you're – that way everything is gonna stack up correctly when you get through. But I'm would talk to Dave about this, but – and first of all, make sure this is gonna do what you wanna do.

Originality evaluation: 'The concept is too similar to what is out there, and you should scrap it'.

The transcript extract shown in Table 24 presents an originality evaluation that is associated with a 'kill' decision for the whole concept, as opposed to recommendations concerning how to rectify the concept through further developments or enhancements (e.g., by changing or adding elements).

Table 24. An extract showing an originality evaluation associated with a 'kill' decision for the whole concept

Speaker	**Discussion**
	(Graduate; Eva; Client review; line 77-83)
Chuck:	This one seems a little far-fetched. I mean, like I – like I said, I appreciate the, uh, I appreciate the out, ya' know, the thinking outside the box, but it's, I mean, maybe we're too – in too much reality.
Eva:	Yeah.

Chuck:	That's why we do these things 'cause this is something I would never think of, but I, I appreciate that you did and, ya' know, this'll probably come to pass someday.
Eva:	Yeah, that's why I give the rating of two stars.
Chuck:	No, that's good.
Eva:	Okay. So I –
Chuck:	Yeah, no, I, I, I definitely appreciate the can. I mean, always push stuff out like that 'cause you never know. You might not have gotten to the last concept if you didn't get to that one so.

4 General discussion

Our analysis aimed to develop a more in-depth theoretical understanding of the evaluative practices that are associated with the 'design critiques' that form a central feature of design education. Such critiques often involve the supportive, critical appraisal of the originality, value and usefulness of the creative design ideas produced by students, with the evaluator being an expert possessing knowledge and competence within the design domain in question. Design critiques typically focus on so-called 'preinventive' structures (e.g., sketches, ideas, and prototypes; see Finke et al., 1992), and trigger further creative exploration that is directed toward improving the idea, although occasionally the critique may lead to an idea being entirely discarded.

Our reported research was predicated on the minimal assumption that there are at least three important high level dimensions to creativity in industrial design practice, that is, originality, functionality and aesthetics, such that a design critique might involve the evaluation of any of these three dimensions of a concept or idea. As we noted in our introduction, it appears that the extant design and creativity literatures have tended to ignore the issue of the 'logics' behind these three different dimensions of creativity and how these logics determine the way in which these dimensions are evaluated in practice. Our conceptual analysis of these dimensions of creativity led to a number of key propositions concerning the way in which these dimensions might trigger particular reasoning strategies and suggestions for design idea progression.

First, we proposed that the three creativity dimensions would diverge in relation to the ontological basis of their associated evaluations, with functionality evaluations standing out from other evaluation types as being based around tests and simulations aimed at assessing whether a design concept meets an objective threshold. In terms of specific, comparative predictions this overarching proposition was partially supported by the observation that functionality evaluation was associated with significantly increased suggestions for testing than aesthetic evaluation. Contrary to expectations, mental simulation was not initially found to be a significant predictor in either the comparison between functionality evaluation and aesthetic evaluation or the comparison between functionality evaluation and originality evaluation. However, a revised analysis that eliminated collinearity effects arising from correlated predictors successfully revealed that mental simulation was significantly more likely to be associated with functionality evaluation than with originality evaluation.

Second, we suggested that the three evaluation types would diverge when it comes to the conception of whether an 'idea' is conceptualized as a 'unit' or as a 'process'. We contended that functionality evaluation and aesthetic evaluation would be essentially process-oriented in that design concepts would be linked to continuous, iterative development. In contrast, we suggested that originality evaluation would be more unit-based or product-oriented, because of the focus on the birth of novel ideas and the choice amongst alternative, competing design 'entities'. Essentially, for originality evaluation our contention was that value is placed on a concept being new, with the greatest levels of value arising when a concept is viewed as being the first of a kind. Our comparative analyses partially corroborated our predictions in that both functionality evaluation and aesthetic evaluation were associated with significantly less 'go/kill' suggestions than originality evaluation. In addition, functionality evaluation (compared to originality evaluation) was associated with significantly fewer suggestions for information search and significantly more suggestions for changing elements of the form or function of the design concept through further development.

Third, we proposed that evaluation types would diverge in terms of the importance of the perception and 'feel' of the design object during the evaluation process. In this respect aesthetic evaluation stands out as being likely to lead to affective judgments based around that 'gut feel' of a concept having 'potential' – but without the evaluator necessarily being able to articulate precisely what this potential might actually mean in terms of how the design should be taken forward through further development. Our proposals in this respect translated into comparative predictions, whereby we expected aesthetic evaluation to be associated with suggestion for the concept to be developed through trial and error to a greater extent than would arise for other evaluation types. This prediction was again partially supported by our comparative analyses, which revealed that aesthetic evaluation was significantly more likely to be associated with suggestions for trial and error behavior.

Overall, then, we believe that our findings attest to the importance of pursuing a detailed, theory-driven assessment of the evaluation behaviors and suggestions for design idea progression that arise in contexts associated with student/expert design critiques. In the present examination of industrial design situations that involved design critiques we found a good level of support for many of our predictions concerning the way in which different dimensions of design creativity trigger different evaluation strategies and idea progression suggestions from expert evaluators. As such, our analysis informs a theoretical understanding of the process of design evaluation as typically arises in an educational context, whilst also having practical implications in terms of alerting expert design evaluators to the nature and consequences of their critical appraisals.

It is interesting to note that while many creative evaluation frameworks (e.g., Amabile, 1982; Csikszentmihalyi, 1990) focus on (gatekeeper) social consensus as the driver of any domain development, what we find when diving into the minds of experienced designers applying different evaluative types is a much more complex picture of creative evaluation. The present analysis suggests that multiple, distinct evaluation logics operate simultaneously, and in predictable ways, in design evaluation. It would therefore seem an oversimplification to claim that creative evaluation within the design process is merely about consensus. Rather, evaluation seems to depend on both social and individual (e.g., hedonic) spheres as well as on different and distinct conceptualizations of what constitutes an idea; it may even relate to distinct ontological assumptions against which concepts are measured. While creative evaluation theorists may at

this point be scratching their heads in trying to bridge the seemingly opposing assumptions in the aforementioned logics, the experienced designers nonetheless seem effortlessly capable of shifting amongst them in actual design practice. More research is needed in order to bring forth the details and further implications of such logics across creative disciplines.

We acknowledge that the specific proportions of evaluation types that we observed in the present analysis may well be unique to industrial design situations. It would therefore be valuable to see further research being undertaken to explore whether the same kinds of evaluation strategies arise in other design domains and occur in similar proportions, whilst also revealing similar kinds of associations with suggestions for design idea progression. We suspect that in more artistic design domains it might be the case that evaluators would be seen to devote proportionally more time and effort to the critique of aesthetic and originality dimensions given the inability to be able to test 'objectively' the functionality of concepts. We note that in relation to the DTRS-10 transcripts, it should be possible to cross-check our findings in relation to the choreography data, which would enable us to explore the extent to which our findings generalize. Although limited time has prohibited this analysis for the present paper we anticipate taking our analysis in this interesting direction in the future.

Acknowledgements

This research project was partly sponsored by a grant from the Strategic Danish Research Counsel to the CIBIS project: Creativity in Blended Interaction Spaces.

References

Adams, R.S. & Siddiqui, J. (2013). *Purdue DTRS 10: Design Review Conversations Database.* XRoads Technical Report TR-01-13, West Lafayette, IN.

Amabile, T. M. (1982). Social psychology of creativity: A consensual assessment technique. Journal of Personality and Social Psychology, 43, 997–1013.

Amabile, T. M. (1996). Creativity in Context: Update to "The Social Psychology of Creativity". Boulder, CO: Westview Press.

Ahmed, S., and Christensen, B. T. (2009). An in situ study of analogical reasoning in novice and experienced design engineers. Journal of Mechanical Design, 131, 111004.1–111004.9.

Ball, L. J., and Christensen, B. T. (2009). Analogical reasoning and mental simulation in design: Two strategies linked to uncertainty resolution. Design Studies, 30, 169–186.

Ball, L. J., Onarheim, B., and Christensen, B. T. (2010). Design requirements, epistemic uncertainty and solution development strategies in software design. Design Studies, 31, 567-589.

Christensen, B. T., and Schunn, C .D. (2007). The relationship of analogical distance to analogical function and pre-inventive structure: The case of engineering design. Memory & Cognition, 35, 29-38.

Christensen, B. T., and Schunn, C. D. (2009). The role and impact of mental simulation in design. Applied Cognitive Psychology, 23, 327–344.

Czikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience. New York: Harper & Row.

Dahl, D. W., and Moreau, P. (2002). The influence and value of analogical thinking during new product design. Journal of Marketing Research, 39, 47-60.

Finke, R. (1990). Creative Imagery: Discoveries and Inventions in Visualization. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Finke, R., Ward, T.B., and Smith, S.M. (1992). Creative Cognition: Theory, Research, and Applications. Cambridge, MA: MIT Press.
Fleiss, J. L. (1981). Balanced incomplete block designs for inter-rater reliability studies. Applied Psychological Measurement, 5, 105-112.
Fleiss, J. L., Levin, B., & Paik, M. C. (1981). The measurement of interrater agreement. Statistical methods for Rates and Proportions, 2, 212-236.
Leder, H., Belke, B., Oeberst, A., & Augustin, D. (2004). A model of aesthetic appreciation and aesthetic judgments. British Journal of Psychology, 95, 489-508.
Mann, L & Tekmen-Araci, Y. (2014). Describing Creativity in Design Across Disciplines. Proceedings of the Design Thinking Research Symposium 10, Purdue University, US, 2014.
McDonnell, J. (2014). Becoming a designer: Some contributions of design reviews. Proceedings of the Design Thinking Research Symposium 10, Purdue University, US, 2014.
Meyer, R. E. (1999). Fifty years of creativity research. In R. J. Sternberg (Ed.), Handbook of Creativity (pp. 449-460). Cambridge: Cambridge University Press.
Moeran, B., and Christensen, B. T. (Eds.) (2013). Exploring Creativity: Evaluative Practices in Innovation, Design and the Arts. Cambridge: Cambridge University Press.
Moldovan, S., Goldenberg, J., and Chattopadhyay, A. (2011). The different roles of product originality and usefulness in generating word-of-mouth. International Journal of Research in Marketing, 28, 109-119.
Nelson, H. G., & Stolterman, E. (2012). The Design Way: Intentional Change in an Unpredictable World. 2nd Edition. Cambridge, MA: MIT Press.
Oak, A. & Lloyd, P. (2014). 'Wait, wait: Dan, your turn': Authority and Assessment in the Design Critique. Proceedings of the Design Thinking Research Symposium 10, Purdue University, US, 2014.
Oh, Y., Ishizaki, S., Gross, M. D., & Yi-Luen Do, E. (2013). A theoretical framework of design critiquing in architecture studios. Design Studies, 34, 302-325.
Richardson, M., & Ball, L. J. (2009). Internal representations, external representations and ergonomics: Towards a theoretical integration. Theoretical Issues in Ergonomics Science, 10, 335-376.
Trickett, S. B., and Trafton, J. G. (2002). The instantiation and use of conceptual simulations in evaluating hypotheses: Movies-in-the-mind in scientific reasoning. In Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society (pp. 878-883). Mahwah, NJ: Erlbaum.
Trickett, S. B., and Trafton, J. G. (2007). "What if": The use of conceptual simulations in scientific reasoning. Cognitive Science, 31, 843-875.
Wiltschnig, S., Christensen, B. T., and Ball. L. J. (2013). Collaborative problem–solution co-evolution in creative design. Design Studies, 34, 515–542.
Yilmaz, S. & Daly, S. R. (2014). Influences of feedback interventions on student ideation practices. Proceedings of the Design Thinking Research Symposium 10, Purdue University, US, 2014.

Biographies

Bo T. Christensen is Associate Professor at the Copenhagen Business School, Denmark. He holds a MSc in Psychology and a PhD in Cognitive Psychology and studies professional creative cognitive processes in a variety of industries using both ethnographic methods, and experimental lab based setups.

Linden J. Ball is Professor of Cognitive Psychology and Dean of the School of Psychology at the University of Central Lancashire, Preston, UK. He holds a BSc degree in Psychology and a PhD in Design Thinking. His research involves laboratory studies of fundamental reasoning and decision making processes and real-world studies of design cognition.