Dimensions of creative evaluation: Distinct design and reasoning strategies for aesthetic, functional and originality judgments

Bo T. Christensen, Copenhagen Business School, Department of Marketing, Solbjerg Plads 3, 2000 Frederiksberg C, Denmark
Linden J. Ball, School of Psychology, University of Central Lancashire, Preston, UK

We examined evaluative reasoning taking place during expert 'design critiques'. We focused on key dimensions of creative evaluation (originality, functionality and aesthetics) and ways in which these dimensions impact reasoning strategies and suggestions offered by experts for how the student could continue. Each dimension was associated with a specific underpinning 'logic' determining how these dimensions were evaluated in practice. Our analysis clarified how these dimensions triggered reasoning strategies such as running mental simulations or making design suggestions, ranging from 'go/kill' decisions to loose recommendations to continue without directional steer. The findings advance our theoretical understanding of evaluation behaviour in design and alert practicing design evaluators to the nature and consequences of their critical appraisals.

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

Evaluative practices are important in all creative industries, where key individuals are invited to assess products 'in-the-making' during initial, creative stages as well as finalised products prior to communicating them to the market (Amabile, 1982; Moeran & Christensen, 2013). Most creative industries have formalised specific roles for domain experts who help advance the initial creative process or who evaluate the final outcome at gates, reviews or screenings. The 'design critique', which is a key feature of design education, is one example of such an evaluative practice, taking the form of a friendly, critical appraisal aimed partly at evaluating the potential, novelty and value of the product in-the-making, but equally importantly serving to catalyse the pursuit of new lines of creative inquiry. The critique presents an opportunity for students to develop their own design values and preferences and to become aware of their own design sensibilities (McDonnell, 2014). In an educational setting the design critique also enables students to reflect upon both the design process and the state of the design, and allows the instructor to reflect on the students' performance (Cardoso, Eris, Badke-Schaub, & Aurisicchio, 2014). Design critiques may play out in

Corresponding author:
Bo T. Christensen
bc.marktg@cbs.dk

www.elsevier.com/locate/destud
0142-694X *Design Studies* **45** (2016) 116–136
http://dx.doi.org/10.1016/j.destud.2015.12.005

many different relationships, from master–apprentice to peer critiques, using various modalities, including speech, gesture and sketching (Oh, Ishizaki, Gross, & Do, 2013). The outcome of design critiques may occasionally be a discarded project, but more frequently they initiate a series of investigations and creative processes aimed at strengthening the project.

The dialogue within design critiques (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 formalised ideas or concepts, and prototypes), in line with the conceptualisation of the creative process offered in the 'Geneplore' model (Finke, 1990; Finke, Ward, & Smith, 1992). This model considers exploratory processes (e.g., contextual shifting and form-before-function reasoning) as inherently 'creative' in nature. This implies that exploratory processes should not be overlooked and that the commonly held belief that creativity primarily concerns *generation* as opposed to *exploration* is mistaken. Indeed, existing design research further underscores the critical role of exploratory enquiry in design problem solving, as emphasised in the theory of problem–solution co-evolution (Dorst & Cross, 2001; Wiltschnig, Christensen, & Ball, 2013). According to this account, processes that alter the problem space such as problem re-framing and contextual shifting co-evolve with changes arising within the solution space. This iterative process of problem–solution co-evolution promotes creative discovery through the building of conceptual 'bridges' between the problem space and the solution space (cf. Cross, 1997).

When considering the potential for creative exploration within design critiques, we note that these critiques usually involve a dedicated and formalised role for the design evaluator, who is presented with a preinventive structure to evaluate and to help advance through an often collaborative process of problem–solution co-evolution. A typical design critique therefore allows for a clear distribution of roles: (1) a designer (or sometimes a design team) who has constructed an initial preinventive structure; and (2) a designer (frequently more experienced) who is exploring, evaluating and helping to develop that preinventive structure. The present research utilises this distribution of roles to examine the different dimensions of creative evaluation in industrial design education and the design strategies employed to enhance creative success. The analysis first and foremost examines how distinct evaluation logics affect the reasoning and progression suggestions of the experienced designer.

In relation to definitions of creativity, a consensus has emerged whereby for a product to be considered 'creative' it must display the properties of *novelty* (or *originality*) and *usefulness* to some domain (e.g., Amabile, 1996; Meyer, 1999; Plucker & Makel, 2010; Sarkar & Chakrabarti, 2011). While novelty is seen as the hallmark of creativity, the arguments for including the usefulness 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. These latter terms, however, are conceptually vague and in need of further specification. In the design domain, Nelson and Stolterman (2012) have provided greater specificity and have listed multiple 'design judgments' that include framing judgments, appearance judgments, quality judgments, compositional judgments and navigational judgments. For our analysis we claim minimally that two high-level and important values in design are *functional value* and *aesthetic value* (e.g., Casakin & Kreitler, 2008; Christiaans, 2002), although we accept that there may be other high-level evaluation dimensions and that our chosen ones may also be separable into finer-grained sub-categories. We further suggest that different value types differ in their underlying evaluation logics, leading, for example, to differences in reasoning strategies.

Below we theorise on the nature of the three dimensions of creativity in industrial design — originality, functionality and aesthetics — and how they may predict differential behaviours for designers and evaluators. The creativity literature has tended to ignore the question of the 'logics' behind these distinct dimensions of creativity and how these logics may relate to the way in which these dimensions are evaluated in practice. In design contexts, for example, design objects can be evaluated from the perspective of different value systems (e.g., functional or aesthetic value), such that the actual 'process' of reasoning about value may take several distinct forms. In the present research we explored 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-occurring — some foregrounded and others backgrounded in concrete evaluative statements), in the present study they are analysed as distinct entities in order to draw out their core differences and respective relations to reasoning strategies.

In this paper we examine the question of how these aforementioned value dimensions are related to distinct reasoning strategies by focusing on design review sessions arising in an educational setting. These design review sessions were drawn from a dataset made available as part of the Tenth Design Thinking Research Symposium (DTRS 10; see Adams, 2015), and involved experienced designers evaluating students' work-in-progress. By focusing exclusively on the reasoning of the experienced designers, we aim to explore how underpinning value dimensions may lead to specific reasoning patterns and suggestions for altering the design. Since the present design sessions relate to a pedagogical context it remains uncertain whether the findings will generalise beyond supervisor/student interactions. Although we do not have any reasons to believe that the patterns identified here will be unique to design education given the generic characteristics of the logics and values proposed, we nevertheless acknowledge up-front the likelihood that: (1) the specific distribution of logics that are valued in different design disciplines may affect the reasoning patterns identified; (2) the pedagogical setting in which the supervisor/student encounter arises could potentially frame the type of evaluative statements offered, for example, by encouraging students to progress with less than optimal designs if the supervisor feels this is as complex a design as the student is capable of developing at the present time; and (3) the stage of design involved (e.g., early versus late) may prove to be an important factor impacting design evaluation (e.g., as argued by Snider, Culley, & Dekoninck, 2013). Further research will thus have to clarify the extent to which the specific patterns shown in the present exploratory study generalise to other design disciplines, settings, and stages of design.

1 The logics of creative dimensions

1.1 Originality evaluation

Most theories of creative evaluation in design make use of an originality or novelty measure (e.g., Amabile, 1983; Charyton, Jagacinski, Merrill, Clifton, & DeDios, 2011; Jansson & Smith, 1991; Shah, Kulkarni, & Vargas-Hernandez, 2000). Evaluations of originality assume that ideas and products exist in objective, temporal reality and that it is possible to analyse their history and development. Value is placed especially on domain-specific originality that may later spark off multiple, fruitful variants in the domain in a germ-like manner (e.g., Shah, Smith, & Vargas-Hernandez, 2003). 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), the other valuing the factual originality for the domain (e.g., as practiced by domain gatekeepers or experts; cf. Amabile, 1982; Csikszentmihalyi, 1990, 2014). The logic of the originality dimension ties it closely to the *birth of ideas* and subsequent 'go/kill' decisions made early in the design process for whole concepts.

1.2 Functional evaluation

As with originality, the evaluation of usefulness or functional value is most often part of the standard dimensions measured when evaluating creativity in design and elsewhere (e.g., Besemer & O'Quinn, 1986; Charyton et al., 2011; Doboli & Umbarkar, 2014; Meyer, 1999; Sarkar & Chakrabarti, 2011), and is argued to relate to goal-satisfaction in design (Shah et al., 2003). Functional evaluation assumes an objective physical reality against which a design concept can be tested. Much functional evaluation involves mentally 'simulating' whether the prescribed requirements are satisfied and whether the design object performs as specified. Mental simulation involves a mental model 'run', which is a change made to a mentally constructed 'model' that enables reasoning about new possible states (Ball & Christensen, 2009; Ball, Onarheim, & Christensen, 2010; Christensen & Schunn, 2009; Wiltschnig et al., 2013). Functional evaluation via mental simulation often focuses on detecting and resolving errors or shortcomings in design elements. While much evaluative design dialogue may revolve around reducing functional uncertainty and turning such uncertainty into approximate answers (Ball & Christensen, 2009; Christensen & Schunn, 2009), ultimately the 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. As such, functional evaluation is fundamentally distinct from socially-oriented consensual agreement as described in much of the creativity evaluation literature (e.g., Amabile, 1996; Csikszentmihalyi, 1990) given the insistence that physical reality remains the ultimate challenge for the functional value of design ideas. Functional evaluation frequently leads to the identification of misbehaving sub-parts that can be improved incrementally (e.g., Chakrabarti & Bligh, 2001). The focus therefore rests on the *life of ideas*, 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.

1.3 Aesthetic evaluation

While it has been claimed that beauty is in the eye of the beholder, research has identified multiple dimensions influencing aesthetic judgments, some relating more clearly to the object itself (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, Belke, Oeberst, & Augustin, 2004; see also Hekkert, 2006). In design, several scholars have argued for the importance of evaluating interaction aesthetics (e.g., Hassenzahl, 2004; Xenakis & Arnellos, 2013). Controversies among art appreciation theorists date back millennia, rendering it unwise to make firm claims about the fundamental nature of aesthetics. Nonetheless, certain qualities of aesthetic judgments in industrial design can be highlighted. 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 be considered higher for aesthetic evaluation compared to originality or functional evaluations. Likewise, aesthetic evaluation may be subject to greater temporal shifts in appreciation in line with the existing social consensus relating to taste or style. Finally, compared to other evaluation types, aesthetic evaluation rests to a larger extent on the affective and cognitive dimensions associated with object perception. The actual perceptual performance seems less important in evaluating originality and functionality, whereas one has to perceive the object with one's own senses to judge aesthetic pleasure. This also implies that judging the aesthetic pleasure of non-perceptual ideas (e.g., designed objects conveyed verbally) is extremely difficult. Materiality matters to aesthetic appreciation, both to the evaluator, but equally so to the creator, where the actual construction and interaction with the designed object is important since the object 'talks back' as it takes shape (Schön, 1983).

2 Propositions and hypotheses

It is important to question what design strategies might be applied in relation to each of the three evaluation types identified. What might we expect in terms of reasoning and suggestions for design idea progression for each evaluation type? Based on the aforementioned differences between the evaluation of originality, aesthetics and functionality, three basic propositions were derived that contextualised the present analysis, as follows:

1. The three types of evaluation diverge in terms of their ontological basis. Functional evaluation stands out given the ability ultimately to test and simulate the capacity for the design to meet objective criteria or requirements. Admittedly, functional evaluation may sometimes be assessed against more subjective criteria, such as the usefulness of the design, but function is most frequently a matter of objectively testable threshold values. As such, functional evaluation should more often lead to suggestions for design experimentation and testing compared to either originality evaluation or aesthetic evaluation. Furthermore, it would be expected that mental simulation of proposed designs would be used as a heuristic strategy or mental shortcut to replace detailed experimental testing.
2. The three evaluation types diverge in relation to what an 'idea' entails in the creative process. In general, creativity theories dissociate in terms of whether ideas are perceived as 'units' (e.g., as in brainstorming, Osborne, 1953, and in blind-variation-selective-retention models of creativity, such as those espoused by Campbell, 1960, and Simonton, 1999) or as 'processes' (e.g., as in stage-gate models, such as that advanced by Cooper, 2001, whereby ideas are viewed as conceptual development over time). 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 functional 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 functional evaluation focus on the life of ideas, and the continual improvement of design through the development of elements by means of additions and changes.

3. Finally, the three evaluation types diverge in terms of the importance during the evaluation process of direct perceptual interaction with the design object. Aesthetic evaluation stands out in this respect, with perceptual interaction with the design object seeming to be especially important for eliciting emotional responses. 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 verbalise 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 given 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.

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

- Comparing aesthetic evaluation to functional 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 functional 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).

We believe it is the first time that the logics behind these three types of design evaluation have been theorised 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 these three evaluation types in actual design practice. As argued in other DTRS 10 papers, the literature on how creativity and creative evaluation varies across disciplines is sparse (Mann & Araci, 2014; Yilmaz & Daly, 2014). Although examining the different proportions of originality evaluation, aesthetic evaluation and functional evaluation across domains is beyond the scope of the present paper, we nevertheless believe that these types of logics may help explain differences in creative evaluation practice, for example, between artistic domains and technical or scientific ones.

3 Methods

The 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, 2015). The data we analysed consisted of 13 supervisor/student pair 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 to 717 for each student and 19 to 470 for each design-critique session. Below we describe the detailed approach adopted for transcript coding (more details on coding may be found in Christensen & Ball, 2014).

3.1 Transcript coding

The transcribed industrial design critiques were independently coded by three students who were unaware of the study's hypotheses. The coders were trained in the analysis of verbal transcripts and were familiarised with the videos and datasets. Each student coded a subset of the data, applying six different codes during five iterations through the datasets. Illustrative codings are presented below, with further examples and qualitative analyses available in (*removed for blind review*).

3.2 Coding of evaluation episodes

We identified all statements uttered by the evaluator (a senior designer) that commented on or evaluated (positively or negatively) a design idea or product. All transcript segments pertaining to evaluative statements below thus derive from the senior designers. The coding excluded comments on the design process, presentation techniques (e.g., PowerPoint visuals) or designers' capabilities unrelated to the designed object. Examples of evaluation statements included: 'that's cool', 'great idea', 'I don't like the x component' and 'this bit might not work'. Following the identification of an evaluation statement, a block of segments relating to this was identified involving descriptions and/or explanations of the design idea (usually uttered before the evaluation statement) and further development or reasoning concerning the evaluation (usually uttered after the evaluation statement). An 'evaluation episode' was then coded, covering the design explanation, the evaluation statement, and the subsequent reasoning/development. Single segments could be coded as evaluation episodes, but they mainly spanned multiple segments.

3.2.1 Coding of evaluation types

All evaluation statements were coded for whether they pertained to aesthetics, function/usage or originality. Evaluations relating to appearance or form were coded as aesthetic evaluations (e.g., as arising in relation to the look, feel or smell of the designed object; Table 1a). Evaluations relating to usage or technical function were coded as functional 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'; Table 1b). 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'; Table 1c).

3.2.2 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), as adapted from Trickett and Trafton (2002; Trickett & Trafton, 2007). Within this coding scheme a mental model 'run' is viewed as a mentally constructed 'model' of a situation, object or system of objects that is grounded in the designer's memory or in their mental modification of design objects that are physically present. Mental simulation enables reasoning 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. Mental simulations are not limited to technical design properties, but also encompass other dynamic situations relating to the designed object such as imagining changes arising from end–user interactions with it or an individual's aesthetic appreciation of some altered component.

Table 1 Transcript extracts showing an example of (a) aesthetic evaluation, (b) functional evaluation, and (c) originality evaluation

Speaker	*Discussion*
(a)	
Gary:	(*Undergraduate*; *Alice*; *2nd review*; *line 121*) This was – save this for another – this one's kinda neat. I really loved how this curved around.
(b)	
Peter:	(*Graduate*; *Mylie*; *Client review*; *line 92*) Ya' know, I love the idea of having accessories that, that can hang from the branches that allow you to customise it and, ya' know, it supports different functionality.
(c)	
Chuck:	(*Graduate*; *Eva*; *Client review*; *line 77*) 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.

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 (Trickett & Trafton, 2007; Richardson & Ball, 2009). This means 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 ending with a final, changed representation (Christensen & Schunn, 2009). These three components of the mental simulation (the initial representation, the simulation run, and the changed representation) are not mutually exclusive and can occur in the same transcript segment, although typically they extend over several segments (see Table 2).

3.2.3 Coding of design idea progression suggestions

These codes captured suggestions for progression of design ideas when an experienced designer evaluated design concepts. Each segment of the transcript was assessed in terms of whether it contained a 'design idea progression suggestion' (DIPS) proffered 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'; Table 3a).
- **Change form or function:** This occurred when a form or function element was added, removed, or changed for a 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'; Table 3b).

3.2.4 Inter-coder reliability checks

To undertake a coding reliability check we selected a set of transcripts involving interactions between a single student and supervisor, which

Table 2 Transcript extract showing an example of a mental simulation

Speaker	*Discussion*
Gary:	(*Undergraduate*; *Adam*; *2nd review*; *line 35–37*) 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.

Table 3 Transcript extracts showing examples of (a) a 'go/kill' DIPS, (b) a 'change form or function' DIPS, (c) a 'test concept' DIPS, (d) a 'search for more information' DIPS, and (e) a 'trial and error' DIPS

- **Test concept:** This arose when the experienced designer suggested testing the concept (e.g., through experimentation or by testing it on users; Table 3c).
- **Search for more information:** This arose when the experienced designer suggested searching for new or additional information for the design (Table 3d).
- **Trial and error**: This 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'; Table 3e).

Speaker	*Discussion*
(a)	
Peter:	(*Graduate*; *Julian*; *Client review*; *line 29*) I think you have other stronger concepts.
(b)	
Chuck:	(*Graduate*; *Mylie*; *Client review*; *line 60*) And, ya' know, maybe you add the fragrance thing in and kinda' take it from there.
(c)	
Gary:	(*Undergraduate*; *Todd*; *Look like*; *line 65*) 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.
(d)	
Peter:	(*Graduate*; *Sydney*; *Client review*; *line 28*) 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.
(e)	
Gary:	(*Undergraduate*; *Alice*; 2nd *review*; *line 177*) 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.

covered three sessions (client review, look like and final review). The transcripts involved 210 segments (approximately 5% of the full dataset). Two individuals coded the transcripts independently, with reliability being estimated using Cohen's Kappa. In case of insufficient reliability, the coding scheme was revised, the coders re-trained, the data re-coded, and new reliability checks conducted. Following the achievement of sufficient reliability all disagreements were resolved through discussion between the coders. All codes displayed satisfactory inter-rater agreement (Table 4), with mental simulation and design idea progression suggestion showing 'fair-to-good' agreement according to the rule-of thumb provided by Fleiss, Levin, and Paik (1981; Fleiss, 1981), whilst the remaining codes showed excellent agreement. Evaluation types showed excellent reliability, illustrating that the forced coding separation into one specific evaluation type (originality, functional, aesthetic) was reliable, and suggesting that the potential for co-occurrence of multiple types in a single segment was not a major problem in the present dataset.

Table 4 Kappa coefficients for inter-coder reliability

Code	*Kappa coefficient*
Mental simulation	.71
Evaluation episodes	.75
Design idea progression suggestion	.68
Evaluation type	.85

4 Results

4.1 Evaluation episodes

We identified 157 unique evaluation episodes, ranging from 1 to 49 segments and averaging 9.9 segments per episode. Evaluation episodes thus made up 36.2% of the transcript segments, which is not surprising given that the essence of design critique is the evaluation of concepts. Following each student across sessions showed that evaluation episodes received by each student ranged from 0 to 32, with an average of 12.1 episodes per student. Of the 157 evaluation episodes, 42% pertained to aesthetic evaluation, 46.5% to functional evaluation and 11.5% to originality evaluation.

4.2 Mental simulation

4.2.1 We identified 113 mental simulations

For each student/evaluator pair, an average of 8.9 mental simulations was carried out, ranging from 0 to 18 mental simulations per pair. Simulation segments occurred much more frequently inside evaluation episodes than outside (Table 5), attesting to the tight coupling between mental simulations and evaluation episodes, $\chi^2(1) = 415.29$, $p < .001$. Only 15 mental simulations did not relate to an evaluation episode in at least one segment.

4.3 Design idea progression suggestions

Across the evaluation episodes there were 153 design idea progression suggestions (DIPS): 45 go/kill DIPS; 67 change 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

Table 5 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

regression analyses. These 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 the five DIPS plus mental simulation, with all predictors coded dichotomously (present vs. absent). Table 6 displays the correlation matrix for predictor variables. Inspection of this correlation matrix indicates that mental simulation and the changing a form or function DIPS were highly correlated ($r\varphi = .700$, $p < .01$), identifying the potential for collinearity confounds in subsequent analyses. These two strategies appear to be highly related because they both involve a changed conception of an initial representation. One way to eliminate collinearity is to remove one of the correlated predictors when conducting subsequent analyses. This approach was taken in all of the analyses that we report below (specific details of predictor removal are provided in relation to each analysis that we report).

To test the hypotheses that we summarised in Section 2, three binary logistic regression models were run for each evaluation type pair, as described below.

4.3.1 H1. Modelling aesthetic versus functional evaluation types

For the aesthetic and functional 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 7 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/.23) times less likely to request trial and error behaviour. These findings support H1a and H1c. To circumvent the potential collinearity confound, the model was run again with the removal of the change form or function DIPS. This change did not yield any differences in the final model, which showed identical predictors and effect sizes to those above.

Table 6 Correlations between predictor variables

	Change form or function	*Test concept*	*Search for information*	*Trial and error*	*Mental simulation*
Go/kill	−.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

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

Table 7 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)*
Step 2	DIPS—test concept	3.16	1.20	6.93	1	.01	23.55
	DIPS—trial and error	−1.47	.67	4.82	1	.03	.23
	Constant	.11	.19	.35	1	.56	1.12

4.3.2 H2. Modelling aesthetic versus 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 (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 to classify correctly, with an overall success rate of 78% (Table 8). 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/.27) times more likely to be associated with the performance of mental simulation. These findings support of H2a and H2b. Again, to control for collinearity confounds, the model was run a second time with the removal of mental simulation. This change did not enable the change form or function DIPS to enter the model, leaving a single variable as a predictor.

4.3.3 H3. Modelling originality versus functional evaluation types

For the originality and functionality evaluation pair, stepwise regression (Wald forward) was carried out, leaving three variables in the final equation (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 to classify correctly, with an overall success rate of 82% (Table 9). The odds ratio indicates that an originality evaluation compared to a functional evaluation is 5.39 times more likely to suggest go/kill decisions (in support of H3c), 15.18 times more likely to suggest searching for more information, and (1/.114) = 8.77 times less likely to suggest changing the form or function of the design concept (in support of H3a).

In the collinearity confound check, removing the change form or function DIPS from the model of functional evaluation and originality evaluation *did* make a difference, with mental simulation now entering the model at the third step, subsequent to the go/kill DIPS and search for information DIPS. A test of the final model versus a model with intercept only was statistically significant, χ^2 (3, N = 85) = 18.87, $p < .001$. The model was able to classify correctly,

Table 8 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	.56	4.59	1	.03	3.28
	Mental simulation	−1.31	.69	3.60	1	.06	.27
	Constant	−1.37	.42	10.83	1	.00	.25

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 functional evaluation was 7.25 (i.e., 1/.138) times less likely to be associated with running mental simulations. In sum, correcting for collinearity by removing a predictor was successful in providing support for the hypothesis that in functional evaluations mental simulations are run more frequently than in originality evaluations, thus corroborating H3b.

Overall, the results indicate how aesthetic judgments frequently lead to suggestions for the designer either to spend time on the concept or to play with it in a trial and error manner without the provision of any further detail as to what is supposed to be developed or changed. In other words, the evaluator leaves it to the student to develop the concept further over a period of time. In contrast, functional evaluation is associated with a mental simulation run to test a concept and progress it via a changed representation. Finally, originality evaluation is associated with a go/kill decision for the whole concept, as opposed to recommendations concerning how to rectify the concept through further enhancements (e.g., by changing or adding elements).

5 General discussion

Our analysis aimed to develop an in-depth understanding of the evaluative practices 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 creative design ideas produced by students, with the evaluator being a knowledgeable and competent domain expert. Our research was predicated on the minimal assumption that there are at least three important high-level dimensions to creativity in industrial

Table 9 Logistic regression (final model) predicting evaluation type (originality vs. functional) from design idea progression suggestions and mental simulation

		B	*S.E.*	*Wald*	*df*	*Sig.*	*Exp(B)*
Step 3	DIPS−go/kill	1.68	.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	.82	7.05	1	.01	.11
	Constant	−1.51	.43	12.40	1	.00	.22

design practice, that is, originality, functionality and aesthetics, such that design critiques might involve the evaluation of any of these dimensions of a concept. As we noted, the extant design and creativity literature appear to have ignored the 'logics' behind these three creativity dimensions and how these logics determine the way in which these dimensions are evaluated. Our conceptual analysis 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 the ontological basis of their associated evaluations, with functional evaluation 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 functional evaluation was associated with significantly increased suggestions for testing compared to aesthetic evaluation. Contrary to expectations, mental simulation was not initially found to be a significant predictor in the comparison between functional evaluation and aesthetic evaluation or the comparison between functional evaluation and originality evaluation. However, a revised analysis eliminating collinearity effects successfully revealed that mental simulation was significantly more likely to be associated with functional evaluation than with originality evaluation.

Second, we suggested that the three evaluation types would diverge in terms of whether an 'idea' is conceptualised as a 'unit' or a 'process'. We contended that functional 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' (i.e., value is placed on a concept being *new*, with the greatest value arising when the concept is the first of a kind). Our comparative analyses partially corroborated our predictions in that both functional evaluation and aesthetic evaluation were associated with significantly less go/kill suggestions than originality evaluation. In addition, functional 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.

Third, we proposed that evaluation types would diverge in terms of the importance of the perceptual 'feel' of the design object during the evaluation process. Aesthetic evaluation stands out as being likely to lead to affective judgments based around the 'gut feel' of a concept having 'potential', but without the evaluator necessarily being able to articulate what this potential might mean for how to take the design forward. Our proposals translated into comparative predictions, whereby we expected aesthetic evaluation to be associated with suggestions 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, with aesthetic evaluation being significantly more likely to be associated with suggestions for trial and error behaviour.

5.1 Implications

Overall, our findings attest to the importance of pursuing a detailed, theory-driven assessment of the evaluation behaviours 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. 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.

While many creative evaluation frameworks (e.g., Amabile, 1982; Csikszentmihalyi, 1990, 2014) 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 evaluation 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 distinct conceptualisations 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, experienced designers nonetheless seem capable of effortlessly shifting amongst them in actual design practice. More research is needed to elucidate the details and implications of such logics across creative disciplines.

Finally, while existing theories of evaluation of creative design tend to view evaluation as an outcome measure in relation to a finalised product (i.e., from process into product, e.g., Sarkar & Chakrabarti, 2011; Shah et al., 2003), the present analysis points in a new directions by showing how the logics of dimensions of creative evaluation feed back into the design process and spark further design development and solution attempts (i.e., from product evaluation to process implications) in predictable patterns of reasoning strategies.

5.2 Limitations and future directions

For design educators the current results highlight the importance of providing feedback on all the different dimensions in design critique sessions. As progression suggestions differ between the creative dimensions it is imperative that students are confronted with all types of evaluation feedback for them to learn the associated creative design skills. Although a creative design is supposed to be simultaneously original, functional and aesthetically pleasing, each of these dimensions relates to different skills and creative processes that students should excel in.

We acknowledge that the specific proportions of evaluation types that we observed may well be unique to industrial design situations. It would therefore be valuable for further research to explore whether the same kinds of evaluation strategies arise in other design domains and occur in similar proportions with similar associations with suggestions for design idea progression. We suspect that in artistic design domains evaluators might be seen to devote proportionally more time and effort to the critique of aesthetic and originality dimensions given the inability to test 'objectively' the functionality of concepts. Further, the educational setting from which the present data are derived could potentially diverge from real-world designing in so far as the experienced designers offering feedback utilise pedagogical approaches that are based on different reasoning strategies from those that arise in normal design practice. We also suggest that future research should try to identify how the progression across design stages affects the reasoning patterns observed here. The present analysis did not have sufficient power to allow for the splitting of reasoning patterns into specific design stages, but further research may well identify how concept development across design stages interacts with different reasoning patterns. Finally, to disentangle the impact of various evaluation types on design reasoning, laboratory-based experimental studies with added cause-effect controls would be valuable to determine whether our present findings can be replicated when eliminating potential confounds.

Acknowledgements

This research was partly sponsored by a grant from the Strategic Danish Research Council to the CIBIS project: Creativity in Blended Interaction Spaces, grant no. 1311-00001B.

References

Adams, R. S. (2015). Design review conversations: the dataset. In R. S. Adams, & J. Siddiqui (Eds.), *Analyzing Design Review Conversations*. West Lafayette, IN: Purdue University Press.

Amabile, T. M. (1982). Social psychology of creativity: a consensual assessment technique. *Journal of Personality and Social Psychology, 43*, 997–1013.
Amabile, T. M. (1983). The social psychology of creativity: a componential conceptualisation. *Journal of Personality and Social Psychology, 45*, 357–376.
Amabile, T. M. (1996). *Creativity in Context: Update to 'The Social Psychology of Creativity'*. Boulder, CO: Westview Press.
Ball, L. J., & 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., & Christensen, B. T. (2010). Design requirements, epistemic uncertainty and solution development strategies in software design. *Design Studies, 31*, 567–589.
Besemer, S., & O'Quinn, K. (1986). Analyzing creative products: refinement and test of a judging instrument. *The Journal of Creative Behavior, 20*, 115–126.
Campbell, D. T. (1960). Blind variation and selective retentions in creative thought as in other knowledge processes. *Psychological Review, 67*, 380–400.
Cardoso, C., Eris, O., Badke-Schaub, P., & Aurisicchio, M. (2014). Question asking in design reviews: how does inquiry facilitate the learning interaction?. In *Paper Presented at Design Thinking Research Symposium* West Lafayette, IN: Purdue University. Paper retrieved from: http://docs.lib.purdue.edu/dtrs/2014/Impact/1/.
Casakin, H., & Kreitler, S. (2008). Correspondences and divergences between teachers and students in the evaluation of design creativity in the design studio. *Environment and Planning B: Planning and Design, 35*, 666–678.
Chakrabarti, A., & Bligh, T. P. (2001). A scheme for functional reasoning in conceptual design. *Design Studies, 22*, 493–517.
Charyton, C., Jagacinski, R. J., Merrill, J. A., Clifton, W., & DeDios, S. (2011). Assessing creativity specific to engineering with the revised creative engineering design assessment. *Journal of Engineering Education, 100*, 778–799.
Christensen, B. T., & Ball, L. J. (2014). Dimensions of creative evaluation: distinct design and reasoning strategies for aesthetic, functional and originality judgments. In *Paper Presented at Design Thinking Research Symposium*. West Lafayette, IN: Purdue University. Paper retrieved from: http://docs.lib.purdue.edu/dtrs/2014/Identity/1/.
Christensen, B. T., & Schunn, C. D. (2009). The role and impact of mental simulation in design. *Applied Cognitive Psychology, 23*, 327–344.
Christiaans, H. H. (2002). Creativity as a design criterion. *Creativity Research Journal, 14*, 41–54.
Cooper, R. G. (2001). *Winning at New Products: Accelerating the Process from Idea to Launch*. New York, NY: Basic Books.
Cross, N. (1997). Descriptive models of creative design: application to an example. *Design Studies, 18*, 427–440.
Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. New York: Harper & Row.
Csikszentmihalyi, M. (2014). *The Systems Model of Creativity*. *The Collected Works of Mihalyi Csikszentmihalyi*. Dordrecht: Springer.
Dahl, D. W., & Moreau, P. (2002). The influence and value of analogical thinking during new product design. *Journal of Marketing Research, 39*, 47–60.
Doboli, A., & Umbarkar, A. (2014). The role of precedents in increasing creativity during iterative design of electronic embedded systems. *Design Studies, 35*, 298–326.
Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem–solution. *Design Studies, 22*, 425–437.

Finke, R. (1990). *Creative Imagery: Discoveries and Inventions in Visualization*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
Finke, R., Ward, T. B., & 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.
Hassenzahl, M. (2004). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction, 19*, 319–349.
Hekkert, P. (2006). Design aesthetics: principles of pleasure in design. *Psychology Science, 48*, 157–172.
Jansson, D. G., & Smith, S. M. (1991). Design fixation. *Design Studies, 12*, 3–11.
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., & Araci, Y. T. (2014). Describing creativity in design across disciplines. In *Paper Presented at Design Thinking Research Symposium*. West Lafayette, IN: Purdue University. Paper retrieved from: http://docs.lib.purdue.edu/dtrs/2014/Multiple/3/.
McDonnell, J. (2014). Becoming a designer: some contributions of design reviews. In *Paper Presented at Design Thinking Research Symposium*. West Lafayette, IN: Purdue University. Paper retrieved from: http://docs.lib.purdue.edu/dtrs/2014/Identity/3/.
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., & Christensen, B. T. (Eds.). (2013). *Exploring Creativity: Evaluative Practices in Innovation, Design and the Arts*. Cambridge: Cambridge University Press.
Moldovan, S., Goldenberg, J., & 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 ed.). Cambridge, MA: MIT Press.
Oh, Y., Ishizaki, S., Gross, M. D., & Do, E. Y.-L. (2013). A theoretical framework of design critiquing in architecture studios. *Design Studies, 34*, 302–325.
Osborne, A. F. (1953). *Applied Imagination: Principles and Procedures of Creative Problem Solving*. New York, NY: Charles Scribener's Sons.
Plucker, J. A., & Makel, M. C. (2010). Assessment of creativity. In J. Kaufman, & R. J. Sternberg (Eds.), *The Cambridge Handbook of Creativity* (pp. 48–73). Cambridge: Cambridge University Press.
Richardson, M., & Ball, L. J. (2009). Internal representations, external representations and ergonomics: towards a theoretical integration. *Theoretical Issues in Ergonomics Science, 10*, 335–376.
Sarkar, P., & Chakrabarti, A. (2011). Assessing design creativity. *Design Studies, 32*, 348–383.
Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. New York, NY: Basic Books.
Shah, J. J., Kulkarni, S. V., & Vargas-Hernandez, N. (2000). Evaluation of idea generation methods for conceptual design: effectiveness metrics and design of experiments. *Journal of Mechanical Design, 122*, 377–384.

Shah, J. J., Smith, S. M., & Vargas-Hernandez, N. (2003). Metrics for measuring ideation effectiveness. *Design Studies, 24*, 111−134.
Simonton, D. K. (1999). Creativity as blind variation and selective retention: is the creative process Darwinian? *Psychological Inquiry, 10*, 309−328.
Snider, C. M., Culley, S. J., & Dekoninck, E. A. (2013). Analysing creative behaviour in the later stage design process. *Design Studies, 34*, 543−574.
Trickett, S. B., & 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., & Trafton, J. G. (2007). "What if": the use of conceptual simulations in scientific reasoning. *Cognitive Science, 31*, 843−875.
Wiltschnig, S., Christensen, B. T., & Ball, L. J. (2013). Collaborative problem−solution co-evolution in creative design. *Design Studies, 34*, 515−542.
Xenakis, I., & Arnellos, A. (2013). The relation between interaction aesthetics and affordances. *Design Studies, 34*, 57−73.
Yilmaz, S., & Daly, S. R. (2014). Influences of feedback interventions on student ideation practices. In *Paper Presented at Design Thinking Research Symposium*. West Lafayette, IN: Purdue University. Paper retrieved from: http://docs.lib.purdue.edu/dtrs/2014/Impact/5/.