Situated emotion and its constructive role in collaborative design: A mixed-method study of experienced designers

Xiao Ge, Larry Leifer and *Linlin Shui*, Center for Design Research, Stanford University, USA

The process of design involves rich emotional experiences from resolving team conflict to breaking impasses to discovering creative insight. To explore emotion in designing we present a mixed-method study of experienced designers. Situated emotion is assessed by triangulating speech acoustics, electrodermal activity, and semantics with contextual inquiry of video data and retrospective self-reporting. We show the value of triangulating multimodal data in dealing with discrepancies amongst different measures. We present results to show how situated emotion mediates design work using multi-level model analysis. To conclude, we discuss how emotion relates to design ability and the role of emotion in designing. Finally, we reflect on the methodological approach we have taken and indicate directions for future emotion research in the process of design.

Keywords: collaborative design, design behaviour, design expertise, design study, emotion

Design thinking, a term that is widely believed to carry the core of design ability, is often criticized for its lack of doing (Micheli, Wilner, Bhatti, Mura, & Beverland, 2019). But there is another missing block, which is design feeling (Coyne, 2005). "We are not thinking machines. We are feeling machines that think", as Antonio Damasio asserted in *Descartes' error* (Damasio, 2006). Emotion is key to understanding designers' ability and behavior. Good designers are known to be able to overcome "emotional blocks" (Maslow, Arnold, in Clancey, 2016; Crilly, 2019) to gain creative insights, which is regarded as a "highly emotional step" by Dorst and Cross (2001). In the creativity literature, Runco (1999) and others have proposed that the emotional discomfort of attempting to reconcile problems would mobilize a creative energy directed at resolving problems through creative expressions and discoveries. Emotion is considered not only functional to the process of creative insight generation, but also key to understanding superior design judgment. Vincenti (1990) argued that good design, especially high-level design, calls for "a great deal of tacit knowledge

Corresponding author:
Corresponding author.Xiao Ge
xiaog@stanford.edu

www.elsevier.com/locate/destud
0142-694X *Design Studies* **75** (2021) 101020
https://doi.org/10.1016/j.destud.2021.101020

related to design feeling". Bucciarelli (1994) eloquently wrote, "one might claim the underlying form of the chair was embedded in the designer's feel for an awl in making a cut". Emotion is also inherent in the social activities of design. For instance, many products come into being out of designers' frustration and dissatisfaction with the status quo (e.g., Sam Farber's OXO Good Grips, Bennett, Peil, & Rosner, 2019). During the design process, a big source of energy and stress in design is working in teams (Kilker, 1999; Dym & Little, 1999; Jung, 2016; Paletz, Schunn, & Kim, 2011). Oftentimes, designers need a fervent emotion, a "daring spirit", to fight for what they think is right (Arnold, in Clancey, 2016).

Most studies of designers, however, have only focused on their behavioral and cognitive characteristics (Casakin & Goldschmidt, 1999; Chiu, 2003; Ahmed & Wallace, 2004; Cross, 2004; Yilmaz & Seifert, 2011; Vallet et al., 2013; Chai, Cen, Ruan, Yang, & Li, 2015). Far less work has empirically examined the role of emotion in design (Sas & Zhang, 2010; Gerber & Carroll, 2012). Understanding how emotion relates to design ability, behavior and process is arguably even more important today, as what face designers are increasingly underdefined problems, unstructured objectives, uncertain processes and unfamiliar stakeholders, which are now the new norms of design in the changing world (Dorst, 2015).

In this paper we present a novel mixed-method approach to identify, characterize and understand situated emotion within experienced designers in the context of collaborative design. We analyze how situated emotion is related to the evolution of design, and then test the hypothesis that intense emotional engagement relates to change of design frame. We discuss how design expertise or ability may mediate emotional responses in situ. Together, our research signifies the important role of emotion in design and strengthens the methodological foundations of design research on emotion.

1 Emotion research in the past

1.1 Measure of emotion in affective science

Mood, affect and emotion research casts a wider net outside design research (e.g., psychology: Csikszentmihalyi, 2013; Davis, 2009; Gino & Ariely, 2012; Ram, Gerstorf, Lindenberger, & Smith, 2011; Lougheed, Brinberg, Ram, & Hollenstein, 2020, management science: Barsade, 2002; Amabile, Barsade, Mueller, & Staw, 2005, learning sciences: Op't Eynde & Turner, 2006; Pekrun, Hall, Goetz, & Perry, 2014). Because how to measure emotion is one of the most vexing problems scholars face (Mauss & Robinson, 2009; Barrett, 2017) and lacks common grounded theories and conceptualizations, we briefly review how emotion has been measured in affective science literature. At the high-level, divergence of emotion research is reflected in what to measure (e.g., discrete states or emotional dimensions) and how to measure (e.g., autonomic, behavioral, or brain states). The emotional dimensionality perspective is recognized to best account for the complexity of emotional behaviors across demographic dimensions and social situations (Barrett, 2017). In this view, emotion is diagrammed by several primitive and universal dimensions (i.e., core affect), such as valence (pleasure - displeasure), arousal (activation - deactivation), dominance, and situational content (Russell, 1980; Barrett, Mesquita, Ochsner, & Gross, 2007).

Emotional arousal has been made easier to measure by technological advances of accessing physiological responding, such as pupil size (Bradley, Miccoli, Escrig, & Lang, 2008), electrodermal activity (EDA, Boucsein, 2012), and speech acoustics (Voigt, Podesva, & Jurafsky, 2014). EDA is the most widely applied psychophysiological response system (Dawson et al., 2017), and can be made less obtrusive using wireless wearables. Another naturalistic setting-friendly approach is speech acoustics. Amongst voice characteristics, voice pitch is the most reliable measure of arousal (Mauss & Robinson, 2009). However, the mapping between the basic physiological processes of emotion and experiences of emotion is still poorly understood (Barrett, 2017).

By contrast, ecological, objective measurements of valence have more limited choices. Eye blink-based startle response magnitude is found to be a robust measure of valence in early emotion research (Mauss & Robinson, 2009), but it only operates in well-defined experimental conditions in the lab. Another reliable indicator of valence is facial behavior (Russell, 1994), which nowadays can be conveniently computed via algorithms (Martinez, 2017). Emotion detection through written text (e.g., twitter, email) is also a widely adopted approach (Kahn, Tobin, Massey, & Anderson, 2007; Calvo & D'Mello, 2010; Vosoughi, Roy, & Aral, 2018), although its validity is rarely examined. One methodological deficit is that participants may censor their true feelings from verbal communication in social settings. Besides, the relation between verbal content and one's emotional experience is rather complex. Similar to physiological measures, these measures of valence also suffer from problems of construct validity (Ram, Brinberg, Pincus, & Conroy, 2017).

Shifting the lens to subjective measures, emotion can be obtained by observer's report (Bartel & Saavedra, 2000; Barsade, 2002; Tsai, Knutson, & Fung, 2006) and self-report. Self-report is typically used in interviews (Sas & Zhang, 2010), reflective journal (Hariharan, 2011), experience sampling (Pychyl, Lee, Thibodeau, & Blunt, 2000; Csikszentmihalyi & Larson, 2014) and surveys where participants are asked to rate their emotions based on a set of scales retrospectively (Barsade, 2002; Todorova, Bear, & Weingart, 2014) or momentarily (Barrett, 1997). In terms of potential limitation, social desirability bias, accessibility and consistency are considered to compromise external validity and study reliability (Ram et al., 2017).

1.2 Study of emotion in design research

The study of designer emotion has a short history. It was initiated in the 2000s and remains relatively marginal. The most widely adopted approach is self-report using qualitative research. In an interview study with expert designers, Sas and Zhang (2010) explored designers' emotional experiences and emotion regulation during different stages of the creative problem-solving process (e.g., incubation stage is characterized by impasse). Hutchinson (2018) explored how graphic designers describe and conceptualize their emotional experiences during ideation with a phenomenography-based interview study. In Gerber and Carroll (2012), an ethnographic study was conducted to inductively explore designers' feelings when their work is facilitated by low-fidelity prototyping. Using a similar approach, Ge and Leifer (2020) explored the emotional learning journey of novice designers going through perplexing experiences within design. Fewer design studies of emotion have applied survey, a mainstream approach outside design research. An example is Hu et al. (2016), where a survey was conducted to examine how engineers' affective states were correlated to levels of task difficulty.

To increase granularity in the self-report assessment, Safin, Dorta, Pierini, Kinayoglu, and Lesage (2016) asked the participants to rate their emotional states using a self-rating tool while self-observing through a video recording. The method allowed the researchers to map designers' emotional patterns during ideation. Jung (2016) applied a similar method to investigate emotional valence using a novel scoring knob. With that, Jung found that a balance of positive and negative affect as well as hostile affect, as measured by participant self-rating, predict design teams' future performance. A potential limitation of this novel method is that it would be impractical to self-rate moment-by-moment experience in ecological settings where the process of design typically lasts much longer than a lab study.

Design studies of emotion have increasingly taken advantage of body language, speech and physiology. Measures based on speech and body language are particularly suitable for early front-end collaborative design, because it is heavily channeled by verbal and physical interactions amongst members of the design team. For instance, Behoora and Tucker (2015) assessed designers' emotional states through body language and machine learning. Another study conducted by Jung (2016) analyzed emotional responses with Specific Affects Coding System, which is a combination of facial muscle movement, speech prosody, verbal content, and body posture. Zhou, Phadnis, and Olechowski (2019) applied facial behavior-based emotion analysis to explore differences of emotion between individual work and group work. In terms of verbal expression-based analysis, Dong, Kleinsmann, and Valkenburg (2009) applied linguistic analysis on design meeting transcripts and found the positive/negative appraisals are associated with different types of knowledge generation and integration. Ewald, Menning, Nicolai, and Weinberg (2019) conducted semantic analysis on transcripts to understand how group emotional valence differs across design thinking stages. Finally, physiological measures such as EDA are being made more accessible outside conventional lab environments in recent years. Villanueva, Campbell, Raikes, Jones, and Putney (2018), for example, collected close-to-real-time EDA data in the wild and used EDA to approximate and compare the emotional experiences of engineering students across different design activities. Rieuf, Bouchard, Meyrueis, and Omhover (2017) used EDA to compare designers' emotional arousals in different design environments. Psychophysiological research of emotion is more widely adopted in interaction design (Balters & Steinert, 2017; Paredes, Ordonez, Ju, & Landay, 2018) but is still nascent in the research of designers' emotions. By contrast, physiology (and neurology) is often used for advancing research of design cognition (Gero & Milovanovic, 2020).

1.3 Research gap in emotion research in design

Apparently, emotion cannot be measured by any single method considered alone (Lang, 1988). Any one measure of emotion has limitations and biases and is likely associated with variance unique to it (Mauss & Robinson, 2009). Partly because of that, different physiological, cognitive, expressive and behavioral measures of emotion have been found to be poorly correlated or uncorrelated at all, showing little support to the emotional response concordance view (Lang, 1988; Cacioppo, Berntson, Larsen, Poehlmann, & Ito, 2000; Mauss, Wilhelm, & Gross, 2004). With more technological convenience made by bodily and linguistic access to emotion, we need to be especially mindful about the choice of emotional measure. Content validity is potentially improved by triangulating observation and interview results, as is typically done in organizational behavior studies (Jick, 1979; Valentine, Nembhard, & Edmondson, 2015). Integrative analysis that combines different emotional measures has proven to provide more valid representations of emotion in some algorithm-augmented computational studies (Kessous, Castellano, & Caridakis, 2010; Poria, Cambria, Bajpai, & Hussain, 2017; D'Mello & Kory, 2015) and multimodal learning analytics research (Blikstein & Worsley, 2016). But it is a rare practice to integrate qualitative and quantitative methods in research of emotion (Hedman, 2014; Schutz, DeCuir-Gunby, & Williams-Johnson, 2016). None of the design studies of emotion have intentionally applied mixed-method approaches. We want to fill this methodological gap.

In addition, most studies reviewed above have not made explicit the links between emotion and design. Instead, they have focused on distinguishing and characterizing designer emotion during a certain design stage or across different stages and situations of design. Some of the studies have further investigated how emotion relates to design outcome (Jung, 2016) and design learning (Dong et al., 2009; Ge & Leifer, 2020). Given the very limited number of emotion studies and our limited understanding of emotion's role in design, we want to further explore their interrelationship and fill this theory gap.

2 Theoretical lens: situated emotion

To study emotion that is relevant and revealing of designers' experiences and processes in their respective ecologically valid settings, we adopt theories that take "situated" views. The theories that situate cognition, intelligence and action in social and physical engagements have a long history since the time of Jean Piaget (Ackermann, 2001) and Kurt Lewin (Lewin, 1999). A few theories that have shaped this paper's perspective are highlighted here.

Bamberger and Schön (1983) use the term 'knowledge-in-action', defined as "the current state of a person's mental constructions", to amplify "the sense of movement and instability" associated to a person's ongoing mental construction as he or she engages with materials at hand. According to both Bamberger (2014) and Schön (1983), a person's knowing is mobile as a result of being situated in and constantly interacting with the changing context instead of being stable, abstract and codified.

Distributed cognition theory is similar in the sense that individuals' thinking is not bounded by the "skin or skull of an individual", but rather being distributed amongst the socio-physical context (Hutchins, 1995). This view is also echoed in Papert's constructionism theory, which places significance on the role of different media in children's knowledge (re)construction (Ackermann, 2001, p. 438).

Despite different takes, these situated views emphasize the mobility of cognition and action, instead of characterizing them as "property of the minds of individuals" (Pea, 1993). Employing such a situated view, scholars (e.g., Hutchins, 1995; Suchman, 1987; Lahlou, 2018) have demonstrated how workers' actions are augmented with, controlled by and channeled through physical and social realities. Design thinking, feeling and doing, in this view, are constructed moment-by-moment through physical, mental and social mediations. The view has been adopted in a few design studies (Kirsh, 2008; Heylighen & Nijs, 2014; Rieuf et al., 2017).

Although emotion has not been explicitly integrated in the theories above, such a sociocultural-constructionist view of emotion is finding traction in affective science (Parkinson, 2012; Mesquita & Boiger, 2014; Barrett, 2017). Mesquita and Boiger (2014) have proposed a sociodynamic model of emotion. In the view of this theory, emotions, despite biological constraints, emerge from the affordance, constraint and reward of sociocultural environments and are functional to the specific sociocultural context. Consistent with situated and distributed views of cognition and action, Mesquita and Boiger have also stressed the inseparableness between emotion and the current socio-cultural environments, the latter being a constituent part of emotion.

Taken together, a situated view of designer emotion assumes socio-physical context plays a key role in the production and management of designers' momentary emotional responses, which in turn influence the evolving cognitive activity and socio-physical interaction. This makes the measurement of situated emotion pointless if ecological validity is not prioritized, and it calls for situated analysis of emotion from multiple analytical angles.

3 Research questions

Employing the theoretical lens above, we applied multiple unobtrusive emotion measures in several dimensions — experiential (retrospective self-report), physiological (speech prosody, EDA), and behavioral (video and semantic analysis) — to enable integrated and repeated examination in context. Given the lack of concordance across emotion measures, we wondered how to make sense of discrepancies, such as when self-report disagree with psychophysiological measure, in order to understand how much the measures are valid, and whether meaningful implications can be drawn about designer behavior. Our first research question is therefore a methodological one (RQ1). The second research question RQ2 explores the interrelation between emotion and design.

RQ1- How can situated emotion be systematically assessed with multiple measures?
RQ2- How are designers' situated emotions related to the evolution of design?

Embedded in RQ2 is a set of sub-questions. Is a designer's situated emotion related to the evolution of their design frame? Here we use Dorst's definition of design frame: it is a standpoint, or a specific perception, from which a problematic situation can be tackled (Schön, 1983; Dorst, 2011). More generally, design frame represents "the way the designer is 'reading' the situation" (Dorst, 2011). We hypothesize there is a positive correlation, given that highly-emotional moments (Dorst & Cross, 2001) signals creative insights, and high-arousal moments of surprise, confusion and curiosity signals reflective behaviors to make new moves, and as a result, change and adjustment of frame (Valkenburg & Dorst, 1998; Schön, 1983). This results in the following hypothesis:

Hypothesis:There is a positive correlation between intense emotional engagement, as suggested by rising emotional arousal and other emotional signals, and a change of design frame.

This hypothesis may especially apply in good, experienced designers, as analyzed in the beginning. Therefore, design ability may mediate the relationship between emotion and design, such that designers with greater adaptive expertise would be more curious to "get stuck" in situations to form new frames that are otherwise perceived as ordinary or uninteresting in the eyes of routine experts or novices. To understand how the interrelation between emotion and design plays out in adaptive expert designers, we explored the questions above for experienced designers, as explained in the next section.

4 Methods

4.1 Participants

10 experienced designers (40% female, 6 European American, 3 Asian American and 1 Latino American) were recruited between June and October of 2019 based on accomplishment and accessibility. They all started as engineers and designers, went through rigorous human-centered design programs from 1970s to 2018 from an engineering department at a U.S. university. None of them collaborated on any design project before. We used Atman's design process conception (Morozov, Kilgore, Yasuhara, & Atman, 2008; Atman, Chimka, Bursic, & Nachtmann, 1999) to assess their confidence, frequency of engagement in design, as well as supportiveness of work environment (Morozov et al., 2008; Atman et al., 2010). Specifically, based on a 0−100 slider scale, the pre-study assessment survey asked the participants, for each design step from need identification to idea implementation, how much they were confident about their design ability, how frequently they were engaged in design, and how supportive their work environment was. The detailed design steps can be found in Atman et al. (1999) and the survey can be found in Morozov et al. (2008). The designers had on average 14.3 years of design work experience. Despite large variance of years of design experience, they were on average highly confident in design ability, received positive work context support and engaged frequently with different kinds of design work. Because of the heterogeneity of engineering and design work, we also retrieved self-identified roles in pre-study survey. Most of the designers were identified with "design". Some designers showed a broader conception of their identification than others. Table 1 shows the designers' design−relevant profiles, where the assessment scores were averaged across design steps.

4.2 Procedures

We created a temporary design studio to imitate a real design environment. The designers worked on an ill-defined design problem for half a day. They were randomly paired to form a design dyad. The dyad form was chosen primarily because team size was constrained by the limited number of EDA wearables and first-person-perspective cameras we had in our lab. Each dyad worked on the given design project for three hours (e.g., 10am to 1pm), on a convenient day for both, based at the temporary design studio. The designers

Table 1 Design work-related participant information

ID[1]	*Design confidence*[2] *(0–100)*	*Supportive work context*[2] *(0–100)*	*Years of design work*	*Work engagement*	*Self-identified role*
Jishin	100	90.71	8	Every design step: > 4 times a day	Design Engineer
Raison	72.86	72.14	2	Gather: a few times a day; Other design steps: occasionally	Research Scientist
Akein	81.57	75.57	27	Prototype, Communicate: frequently; Other design steps: occasionally	Product Designer, Researcher
Emmo	78.57	68.57	3	Gather, Evaluate, Communicate: frequently; Other design steps: occasionally	Engineer, Designer, Product Manager, Researcher, Entrepreneur
River	80.00	80.00	40	Define, Gather, Ideate, Communicate: a few times a day; Other design steps: occasionally	Design Innovator, Product Designer, Consultant, Educator, Researcher, Magician
Phi	97.57	75.86	8	Ideate: > 4 times a day; Other design steps: frequently	Engineering Professional
Jurian	75.71	69.29	6	Every design step: a few times a day	Engineering Designer, Researcher, Research Scientist
Narra	95.71	100	40	Every design step: frequently	Designer, Innovator, Designer, Consultant, Educator
Maanav	86.43	87.86	4	Prototype, Communicate: a few times a day; Ideate, Evaluate: frequently; Other design steps: occasionally	Engineer, Designer, Innovator, Product Manager, Consultant, Educator, Research Scientist
Creta	94.71	70.29	5	Define, Gather, Ideate, Communicate: a few times a day; Other design steps: occasionally	Engineer, Designer, Innovator, Consultant, Research Scientist, CEO
Average	86.31	79.03	14.3	N/A	N/A

Note: 1. Anonymous labels are used to protect participant privacy. 2. Results are averaged across all design steps as described in Atman et al. (1999).

were physically unconstrained from the design studio during the study. Outside the studio is a home-style kitchen and a common area. The design studio is 1-minute walking distance from some residences of families with small children. Before the study began, each designer signed consent forms, completed demographics surveys, put on measuring instruments, and were informed about the design task. Given a set of materials, the designers were asked to work on an early front-end design task, to create solutions to radically improve the dining experiences of families with small children.

The design prompt gave no constraints on what kind of solution to deliver. Most design tasks in past research offered participants a defined solution scope, such as a new trash system (Dorst, 2015), a carrying/fastening device to carry a backpack on a mountain bicycle (Cross, Dorst, & Christiaans, 1996; Cross & Cross, 1998), a playground (Atman et al., 1999), or redesigning an object (Cannon, 2018).

The participants were given a few materials, including pre-collected user videos of eating scenarios, an instruction sheet about the task and the deliverables, a news article related to parental relationships, a map of the design studio's surroundings, and prototyping materials. Two hours into the study, some users — two families with their child(ren) — came in for user testing. We asked the users beforehand not to make only positive remarks, but to provide honest, critical and helpful feedback. By the end, each dyad delivered a pitch, as if to a client, with some prototypes in hand. An example of the process the participants went through is shown in Figure 2.

At the end of their projects, the design dyads came up with various concepts and prototypes. Some examples are:

- A service design idea to create shared caring for young parents living far away from their support system.
- A pre-meal meditation product to allow parents to be centered and connected with their children.
- Special food-and-utensil set for young children, in which each eating tool is designed to be used together with a specific kind of food, to bridge "how to eat" with "what to eat".
- A set of eating tools to engage kids during dining, called "distraction as food".
- A real cooking station, with "mixing bowl" and "easy-bake oven", to engage kids in the last-step cooking preparation.
- A "food-on-demand" system to allow kids to eat on their timing.

4.3 Measures

We employed multiple observational channels to account for uncontrolled variables in the study setting. Fixed video cameras were located at the four corners of the studio. Each participant wore a miniature eye-level video camera (AXON Flex 2[1]), to capture how situations were lived from the perspective of the participant (Lahlou, 2011; Lahlou, 2018), as well as keep track of designers' activity outside the studio, as shown in Figure 1 (b) and (d).

4.3.1 Skin Conductance Response (SCR) frequency

To capture moment-by-moment emotional arousal without interfering with designers' work, each participant wore the Empatica E4 wristband[2] on the dominant hand which derived the EDA[3] data. EDA was processed with Continuous Decomposition Analysis (CDA) in Ledalab (v3.4.9) with Matlab software (Benedek & Kaernbach, 2010). Skin Conductance Response (SCR) was exported using a minimum amplitude criterion of 0.05 μS in consideration of the uncontrolled study setting (Boucsein, 2012; Caruelle, Gustafsson, Shams, & Lervik-Olsen, 2019), accounting for a 2-second latency (Dawson, Schell, & Filion, 2017). We used SCR frequency to categorize the data into

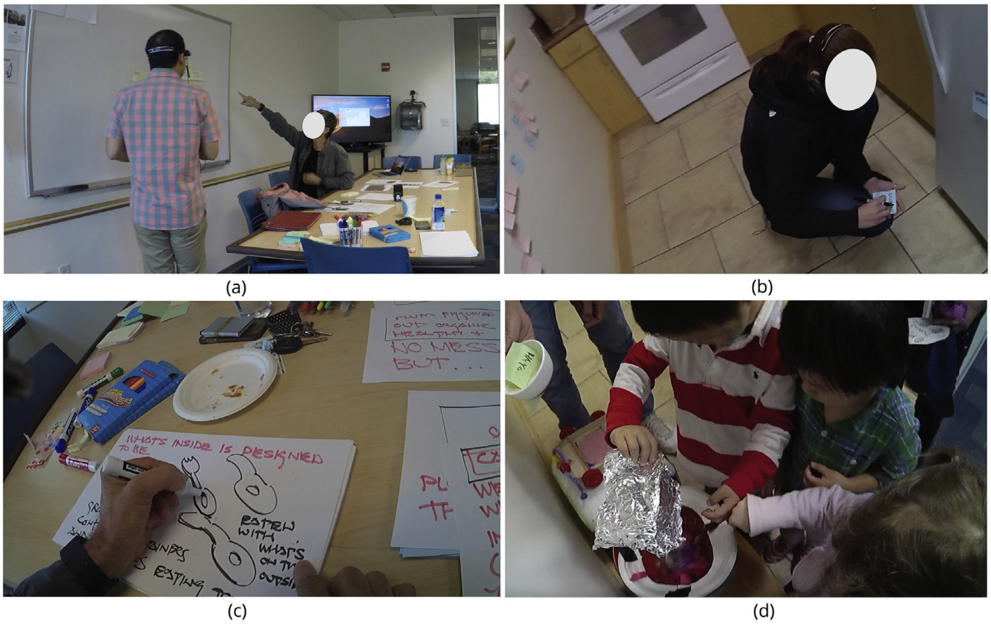

Figure 1 Examples of design work in the study. (a) Two designers mind-mapping the problem space for the design project; (b) In a kitchen outside the design studio, a designer was sketching ideas about refrigerator-based kids-friendly water refill, viewed from the first-person perspective of the collaborating designer; (c) A designer sketching an concept of a utensil set that was customized for the food to be eaten; (d) A dyad of designers testing their prototypes of "mixing bowl" and "easy-bake oven" with three kids and their parents through role playing. Their idea was to make meal experience more fun by engaging children in the last-step cooking preparation

low, medium, and high levels (Boucsein, 2012; Dawson et al., 2017). According to the standard, a frequency of 1–3 peaks/min (ppm) would occur at rest, and as frequency increases with the arousal level, values higher than 20 ppm was interpreted as high arousal, while anything in between was labelled as medium. Signals less than 1 ppm were discarded for the purpose of data cleanup. EDA data from designer Phi was missing due to loose device.

4.3.2 Vocal pitch

The other situated arousal measure was vocal pitch. Audio data was first acquired from the miniature subjective camera, manually cleaned up for speaker diarization and partitioned into utterances with ELAN software (Aguera, Jerbi, Caclin, & Bertrand, 2011). Laughing, whistling, sneezing and other sounds were separated out, because of their different vocal characteristics, but they were included in the overall analysis. Overlapping talks were not included for individual-level analysis. Only voiced speech was used (Dietrich, Enos, & Sen, 2019). We treated each cleaned-up utterance as a unit and extracted vocal pitch with PRAAT software (Boersma, 2006). The standard pitch range (male: 30 to 450 Hz; female: 75 to 600 Hz) was adopted except for some male participants who had rather high pitch. To deal with

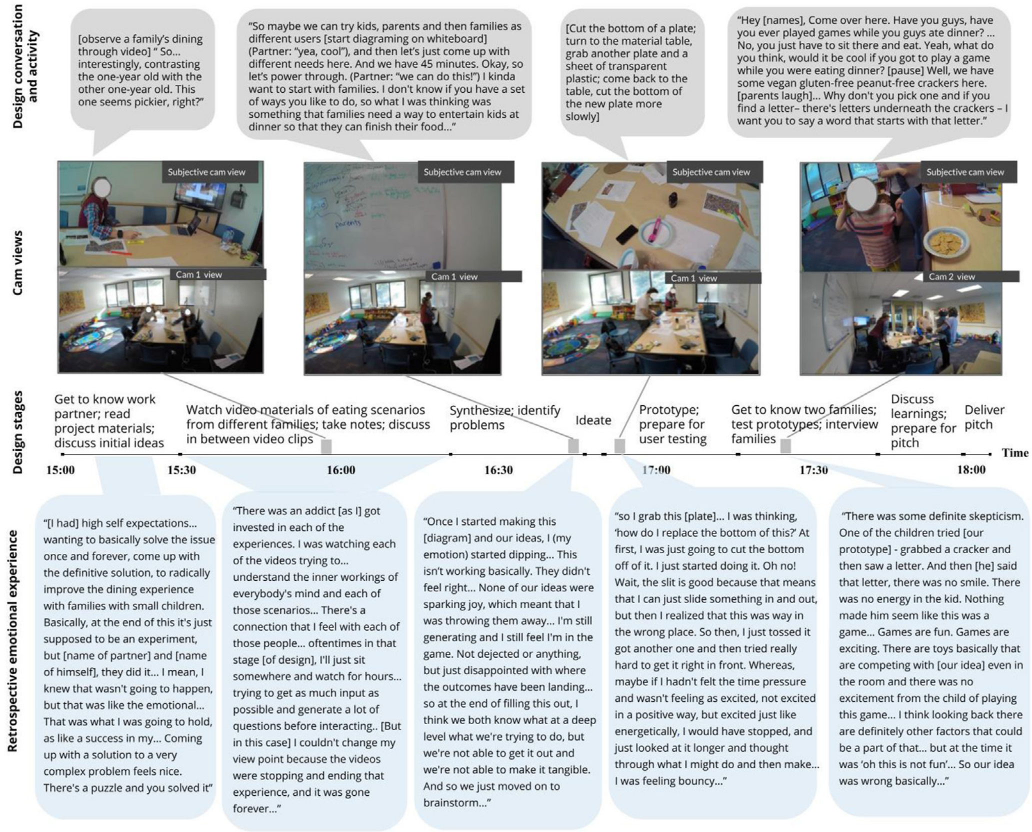

Figure 2 An example of the design process and the qualitative data collected during the study. The example shows time-stamped design conversation, first-person-perspective/subjective camera view and retrospective emotional experience from one of the two collaborating designers. The example gives a tangible view of the kind of design processes and experiences and outcomes from the study

creaky voice problems in some male participants, we broadened the pitch range, which does not affect data output (Dietrich et al., 2019). Remaining creaky voices were detected and removed with the Covarep program in Matlab (Degottex, Kane, Drugman, Raitio, & Scherer, 2014). The final data of vocal pitch contains missing data of the final design phases for dyad 1, due to large background noise and possible device malfunction. To allow for between-person comparison, we re-scaled vocal pitch to z-score, or standard deviations above and below each designer's average vocal pitch (Dietrich et al., 2019).

4.3.3 Retrospective emotional experience

To retrieve emotional experiences from the designers' perspective, within a few days after the design task, the designers reported their experiences retrospectively in a 2-hour one-on-one semi-structured interview with the first author. We followed an interview protocol that focused on asking open-ended questions about their emotional experiences during the study. Large prints of their own camera views, artifacts produced during the design task, video clips and other facilitating tools were used when appropriate to facilitate the unfolding of internal experiences. Emotional valence (positive - negative) and arousal (high - low) as described and visualized by the designers were directly used in analysis (e.g., Figure 3). The reported emotional states and related reflections were used for contextual analysis. Not all designers provided emotional arousal for the whole timespan of the design task. Instead, they talked about it for pronounced events. Some designers were also self-contradictory when they described emotions. For instance, one designer reflected that there were no negative emotions earlier in the interview, but later revealed that he was angry at a user's behavior. In consideration of the lack of full details and inconsistency, we primarily used pronounced experiences as selectively detailed by the designers for interpreting physiological data and video data.

4.3.4 Other emotion measures

Apart from vocal pitch, vocal intensity (i.e., speech loudness) and laughter patterns were extracted for explorative analysis of emotion. Transcripts with reference to time were analyzed by the emotional lexicon collected, manually curated and evaluated by the National Research Council of Canada (NRC) (Mohammad & Turney, 2013). We used NRC lexicon to perform semantic analysis and render scores of emotional valence and arousal for design conversations.

4.3.5 Change of design frame

To examine design frame and how it is adjusted and how it expands and changes over time, we used the "Probability of Acute Change" (PAC) metric (Jahng, Wood, & Trull, 2008) on the time-ordered transcripts to calculate the likelihood of forming or using new words at a certain time of point in design, and then we used this measure to approximate the degree of change in each designer's design frame. PAC represents the likelihood of acute changes across time, defined by the number of acute changes divided by the total number of changes, which in our context is the acute increase of new words. We calculated the increase of new-word count for each dyad-specific design phase using the Python Natural Language Toolkit (Bird, Klein, & Loper, 2009), and whether an acute change of new-word count occurs or not is defined by a cut point. The cut point was chosen by sensitivity analysis (Jahng et al., 2008). Put another way, acute increase of new words in a designer's verbal expression, as compared with their previous use of words, indicates a change of design frame.

4.3.6 Other measures

Retrospective self-report was also used to retrieve the designers' general beliefs about emotion, behavior and ability both within and outside the current study. The multi-angle video data was used for contextualization and triangulation

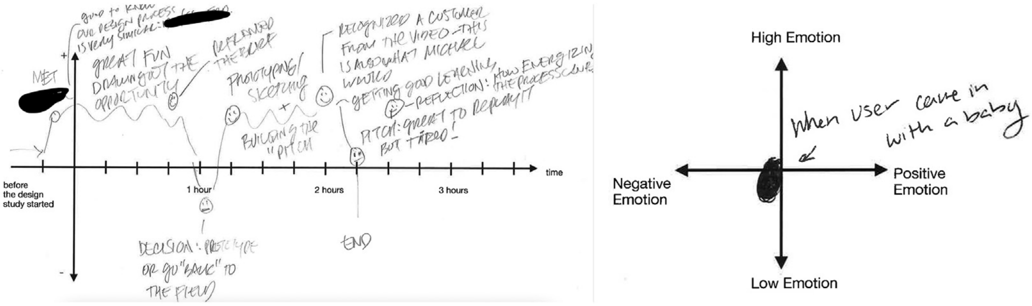

Figure 3 Participant drawings of emotional valence map (left) and emotional valence and arousal matrices (right) were used to facilitate retrospective interview

with the above objective measures. For statistical analysis of the correlation between intense emotional engagement and change of design frame, we included moderating variables such as a time variable and a dummy variable of social interaction (1: user interaction; 0: no user interaction).

4.4 Data analysis

For RQ1, we first analyzed how well different measures converge; with that understanding, we analyzed peculiarities of different measures with references to the specific situation and designers' experiences for more comprehensive understanding of salient events (Xue & Desmet, 2019). Matching scores were derived by comparison between retrospective self-report and other measures, specifically for subjectively intense emotional experiences (Reisenzein, 2000). Vocal pitches that are 1.5 standard deviations above the local average (i.e., z-score) are thereafter called high-pitch incidents, or momentary high-emotion engagements, and are regarded as pointers to pronounced episodes of emotional responses to novel situations (e.g., surprise, curiosity, and confusion).

For RQ2, we first conducted a mixed-method analysis of emotion in the process of design. We then tested the hypothesis with a multilevel model method (Grimm, Ram, & Estabrook, 2016) to analyze the intra-individual time covariation between intense emotional arousal and change of design frame (fixed effect), as well as between-person variances (random effect). For this statistical analysis of design frame and its change, we cleaned up the data further, to exclude intermediate interactions indirectly related to the design project, such as taking a break. Because vocal pitch, vocal pitch variability (i.e., standard deviation of pitch) and probability of high-pitch incidents are highly correlated (correlation coefficients range from 0.6 to 0.8), we present detailed results for vocal pitch variability only. The time variable is included to control for possible time dependencies of the tested correlation, such as design phase-related reasons for rising vocal pitch or change of design frame. We also used lag 1 autoregression (Bolger & Laurenceau, 2013) to further account for any auto-regression time dependencies. To further answer RQ2, we analyzed the role of design expertise in the interrelation between emotion and design.

5 Results

To answer RQ1 (How to systematically assess situated emotion with multiple measures?), we first describe the designers' emotions at the high level with basic results (section 5.1), followed by comparing different measures to evaluate discordance (section 5.2). We exemplify mixed-method approach with the data of two designers (section 5.3), which also provides qualitative evidence for RQ2 (How is designers' situated emotion related to the evolution of design?). To corroborate the finding about the link between emotion and design, we show the results from the multilevel model analysis (section 5.4). In the sections below, quotation marks are used for direct quotes from the designers. High-pitch incidents are quoted in *italics*. All quotes are from conversations that occurred during the study unless noted.

5.1 Descriptive results of physiological measures and semantic analysis

This section describes the designers' emotions at the high level with basic results. As shown in Table 2, the average SCR frequency is 20.71 ppm (SD: 12.62 ppm), which falls into medium-to-high arousal range according to the previously introduced standard, suggesting the designers had medium-to-high engagement on average. There was on average 1388 observations of vocal pitch and vocal intensity during the design project, with average vocal pitch: 167.02 Hz (SD: 51.08 Hz). Female designers used much higher vocal pitch in general than the male designers, but vocal intensities (i.e., speech loudness) were at similar levels. Figure 4 shows diverse vocal behaviors across designers, with some designers having much lower vocal pitch variability (i.e., smaller changes of vocal pitch) on average than the other designers. On the other hand, within each designer, vocal pitch variability fluctuated a lot with time. At the dyad level, some dyads were more vocally balanced between the two design members in terms of vocal pitch and its variability (e.g., dyad 4), as compared with other dyads. Figure 4 also shows that momentary high-emotion engagements, or high-pitch incidents, scattered across time, suggesting the designers were stimulated throughout the design phases.

Semantic analysis shows that vocal pitch is related to emotional words expressed in verbal content, yet there are not any consistent patterns across designers. For instance, River's vocal pitch covaries with high-arousal verbal contents ($p = 0.037$), and with both positive expression ($p = 0.005$) and negative expression ($p = 0.038$). This indicates River used more diverse emotional words − highly-arousal, positive and negative words − when he raised voice.

Table 2 Descriptive results of vocal pitch (non-standardized) and SCR frequency. Consistent with gender difference, the female designers spoke with much higher mean and variance of vocal pitch than the male designers, but vocal intensities were at similar levels

Dyad	*ID*	*# of vocal observations*	*Vocal pitch (Hertz)*		*Vocal Intensity (dB)*		*SCR frequency (ppm)*	
			Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*
1	Phi	1248	196.26	60.98	65.88	5.98	–[1]	–
	Jishin	857	133.59	33.99	65.69	6.35	15.44	15.5
2	River	1360	157.59	54.73	64.24	6.09	30.37	17.12
	Akein	903	103.81	35.02	67.24	5.91	13.69	10.88
3	Narra	1537	118.29	30.54	68.98	6.85	25.48	13.94
	Jurian	1401	117.52	35.79	68.88	4.91	6.85	7.68
4	Maanav	1597	255.03	84.03	71.77	6.39	23.67	14.51
	Creta	1590	273.3	84.76	73.67	6.00	28.77	15.07
5	Raison	1781	106.32	32.70	71.30	5.68	25.96	13.37
	Emmo	1605	208.47	58.25	67.75	5.55	16.19	15.04
Average		1388	167.02	51.08	68.75	6.68	20.71	12.62

Note: 1. EDA data from designer Phi was missing due to loose device.

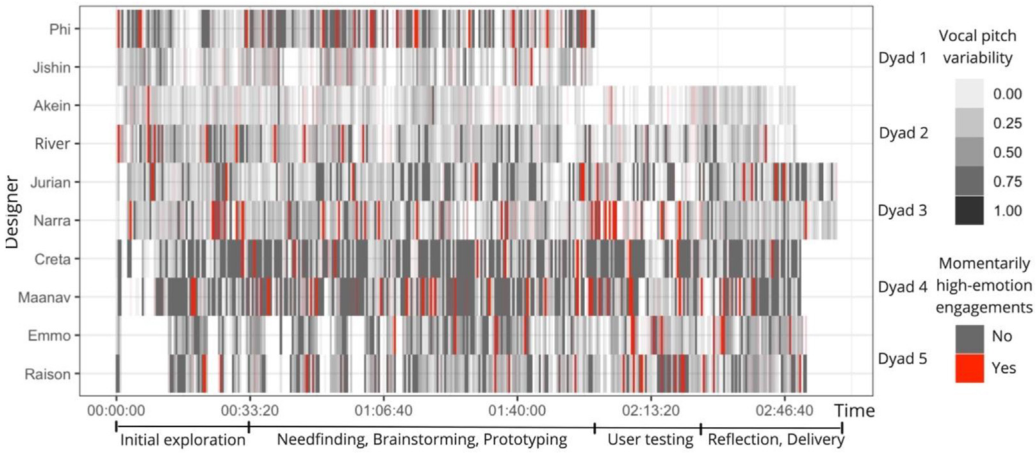

Figure 4 Vocal pitch metrics within and across designers. Momentary high-emotion engagements, or high-pitch incidents, as approximated by z-scores larger than 1.5, spread across all stages. Vocal pitch variability shows between-designer and between-dyad difference. Time is roughly segmented into a few design phases based on the overall design process. White space suggests missing vocal pitch data, mostly caused by the designer not talking during the time. Dyad 1 missed data in the final design phases

Jurian's pitch covaries with positive expression ($p = 0.006$), showing Jurians's high-tone talks were consistently characterized by positive word usage. Creta's pitch correlates with negative emotional word frequency per utterance ($p = 0.052$). In other words, Creta used more negative words when she raised her voice. As for Maanav, her high-tone talking had high-arousal, positive verbal contents, with positive correlation found between pitch and positive expression ($p < 0.001$), and between pitch and high-arousal verbal content

($p < 0.001$). None of the four designers shared the same pattern. The lack of converging semantic patterns across designers implies idiosyncratic semantic covariations with vocal pitch.

5.2 Evaluation of discordance amongst emotion measures

Although vocal pitch is found to be related to verbal content, there is a lack of within-person concordance between the arousal measures of vocal pitch and SCR frequency, and the discordance is also observed when compared with retrospective emotional experience. The generally observed discordance reflects that different measures capture different aspects of emotional arousal.

Retrospective emotional experience gives special access to symbolic values associated with a physiological or behavioral emotional response and reflects sociocultural complexities which may not be able to be translated into a singular value of emotional arousal, as in SCR frequency and vocal pitch. In total, 73.3% of the most pronounced emotional experiences, as reported by the designers, find validation in the respective SCR frequency graphs, and only 40% of them co-occur with rising or dropping pitch average. As an example, to show the associated complexities, River had mixed and contradicting emotional arousal as he entered the design task. River reported he was very stressful (i.e., high-arousal) earlier in the process because of working with an unfamiliar partner; but he was simultaneously neutral and relaxed (i.e., low-arousal) about the project itself.

Such richness of emotion is obscured in physiological measures. SCR frequency generally reflected the designers' perceived level of energy or stress, but it was not able to deal with nuances and contradictory experience as exemplified in this case of River, instead, it has only captured River's great stress of teamwork, as shown in Figure 6a.

Long experiences of stress and energy were often regulated in vocal expressions. In general, design talks were well-regulated as reflected in the narrow distribution of vocal pitch, where most data points gravitate towards the mean. This partly explains why phase-by-phase covariation between mean SCR frequency and mean vocal pitch is only found 2 of the 9 designers: Jurian ($r = 0.44$, $t(32) = 2.78$, $p = 0.009$) and Raison ($r = 0.52$, $t(19) = 2.66$, $p = 0.015$).

Although most vocal expressions were regulated, vocal pitch is sensitive to momentary high-emotion engagements resulted from novel external stimuli. This is reflected in its right-skewed distribution with lots of high vocal-pitch utterances that were stimulated by novel situations. We will show examples of high-pitch incidents in the next section. These momentary high-emotion engagements tend to be swift during a long activity and were less likely to be

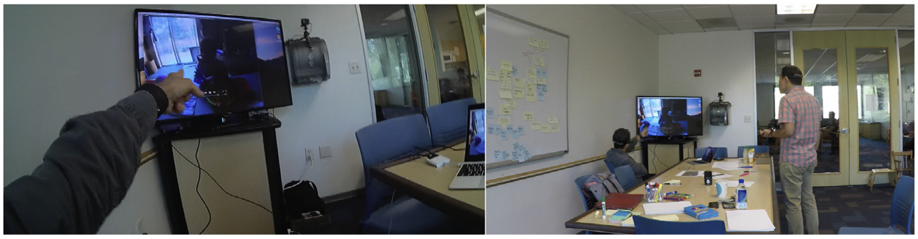

*Figure 5 Designer River's view (left) and a third-person camera view (right) of a high-emotion engagement moment for River. Left: River extends his left arm and points at the TV, saying "*But I'm realizing it's a baby bottle!*" (z-score: 1.53). Right: Akein (standing) observing user behaviors as River (sitting) talks*

voluntarily reported in retrospect despite that they can be quite emotional in situ. This partly explains the low concordance between vocal pitch and retrospective emotional experience.

5.3 *Mixed-method analysis of situated emotion*

Understanding the different qualities of the emotion measures, we can make meaningful situated analysis by capitalizing on discordance and the idiosyncratic contexts the designers were in. Consider River, who had the highest SCR frequency in terms of mean and variance. We will focus on analyzing River's situated emotion, and further make sense of it by comparing it with the situated emotion of Jurian who had the lowest SCR frequency of all the designers in the study.

5.3.1 *Situated emotion of the designer River*

River perceived design as a process of "going along where your energy is". Self-regarded as a divergent thinker, the messy and uneconomic process of exploring *what the design problem really is* fuels him with the necessary energy and motivation to work. Although he was an experienced product designer for children, and had done a lot of work in the design space of food, he allowed himself to be intrigued every time he read or watched a new piece of information. User material from the parents' quotes to video scenes of children at eating was his sandbox to play. His verbal responses to them were also characterized by diverse emotional words of all kinds, as shown by the semantic result. Consider this following 2-minute interaction River had as mediated by a video scene. It marks a time when River reframed the design problem from the tension between emerging roles of a new parent to the problematic positioning of children products. Within the short time, River was channeled to have several high-pitch incidents of surprise, confusion and curiosity. These emotional engagements are also evidenced by his verbal expressions and body language.

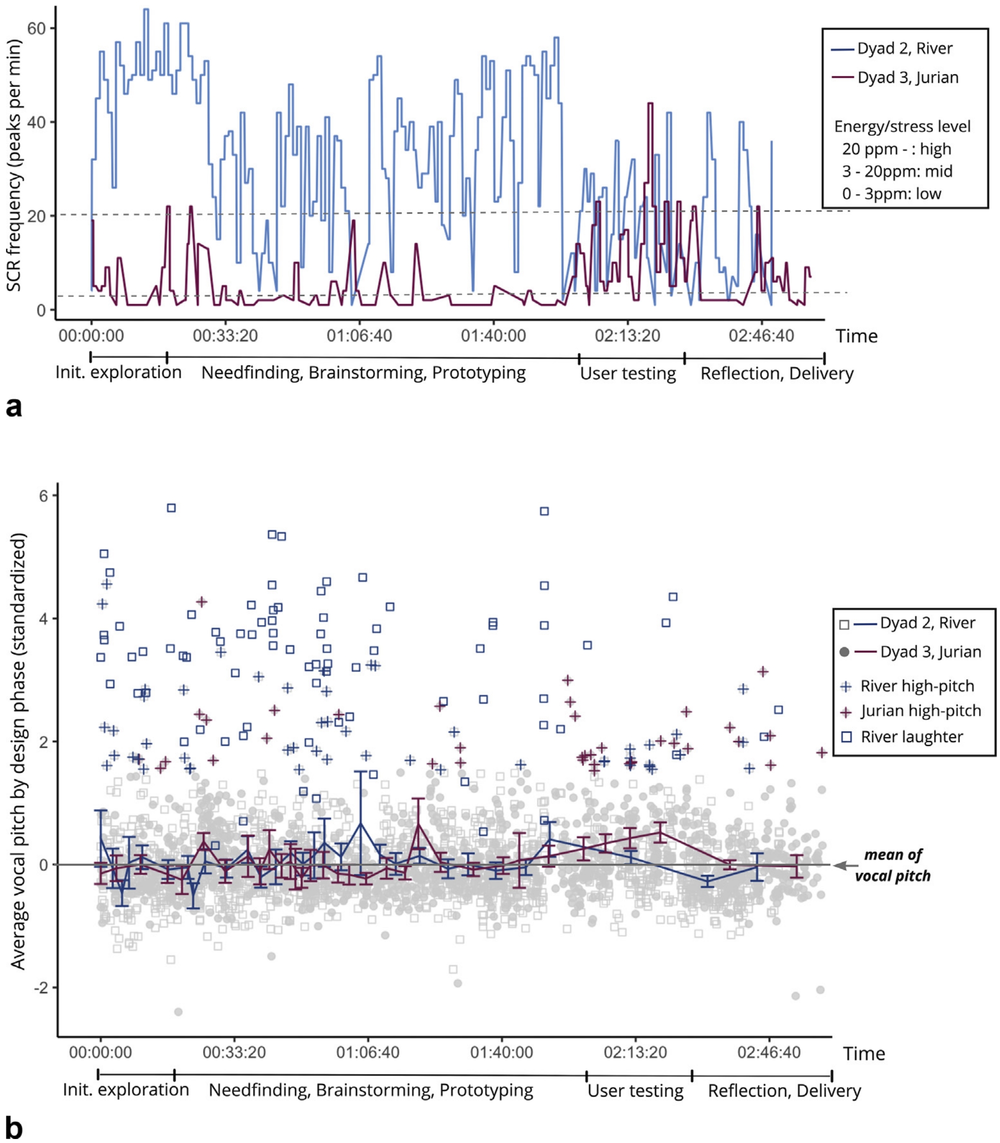

Figure 6 a. Within-person fluctuations of SCR frequency across time, from designer River and Jurian. Dividing lines of arousal levels are based on the standard (Boucsein, 2012; Dawson et al., 2017). River's arousal in the user-testing phase and thereafter are low relative to his average SCR frequency. In comparison, Jurian's SCR frequency peaked in the same phase, where he had a stressful experience with unexpected user-testing situation. b. Within-person fluctuations of average vocal pitch (with error bars of 95% confidence interval). As exemplified by the cases of River and Jurian, between-person differences of vocal pitch are salient in some particular design phases. River's laughter pattern is high-lighted, showing a decrease of laughing behavior in the second half of the project. Comparing the high-pitch incidents between River and Jurian, River had way more high pitches in the first half than the second half of the project, whereas Jurian was equally stimulated throughout

After watching the user video in silence for a while, River eagerly shared his observation with his partner Akein that adult products were scaled down for kids in softer materials; "*but it isn't clear*" (z-score: 1.89), he paused, and did not finish the sentence before continuing, "we've seen these modifications in adult world of plates and utensils. So, within the culture we are trying to kidify it enough that (unfinished and pause) ... It isn't clear it's helping the kids. And I think the parents go, this looks like it is for kids."

He continued with a related observation from past experiences that some adult bottle designs were actually influenced by baby products: "it literally went the other way around!" As the video played along, he turned his attention back and forth between the TV screen and Akein. "*But I'm realizing it's a baby bottle!*" (z-score: 1.53), he excitedly pointed to the screen (Figure 5); "But that…", River quickly brought his hand back to hold his chin, and then stretched out to gesture, contradicting his earlier comment: "*And it works for her! It works perfectly. She had no problem with that!*" (z-score: 1.66).

While Akein acknowledged River's observation, River kept watching the video, holding a fist by his chin, asking in a deeper and lower voice: "So are we looking at this the wrong way, that we wanted to have …", he briefly paused, turned to Akein, and increased his volume, taking a lighter tone, "kids become adults versus learn from the kids and modify our adult world (laughter)! That might not be hard to sell."

Another big source of energy for River was who to play with in the sandbox. It was very important for him to have a work partner of "shared unspoken process" with him. In his reflection, he emphasized that "figuring out the partner was the most important piece of getting through [the design process]". As a result, the temporary teamwork with a stranger shaped the great stress River had at the beginning of the design task. This stress dipped through a positive social mediation in the first half of the project. Every time River learnt a bit more about Akein, he had, as he reflected, an "oh, phew!" experience of delight and relief. For instance, early in the process when the dyad was crystalizing their ideas using a design tool, as his partner Akein commented on their shared design training regarding the design tool, River was positively surprised and joked: "*Okay, checked! Extra credit! We are done, we are out of here!*" (z-score: 1.83). This was followed with both designers in laughter.

Together, the energizing socio-physical interaction intertwined with the stress River had about teamwork, allowing us to make sense of the very high SCR frequency (Figure 6a) and the densely-populated high-pitch incidents (Figure 6b) in the early phases.

The socio-physical stimulations channeled a positive energy for River, until 1.5 h into the design task, when he had a major conflict with Akein about the project direction. River regarded it in the interview as "the high point of perplexity". At this point River suggested talking to some actual users to validate their design assumptions before moving onto ideation and prototyping.

The assumption they had converged upon was around the combined observation of kids struggling to stab and scoop the egg with forks and frustrated parents trying to control the mess associated with their short-attention-span kids. What River regarded as an "assumption" was a temporary conclusion River and Akein had arrived. And that was: families struggle in the big gulf between what kids eat (e.g., baby food or family food) and how kids eat (e.g., kidified utensil or adult product), because companies that work on what kids eat (e.g., Plum's space food) and companies that focus on how kids eat (e.g., Rocketship-shaped utensils) operate separately in their own territories and do not work together. River and Akein believed there was an opportunity to design some products to bridge the gulf so that "kids can enter more easily the family circle without the stress of 'I'm failing around not being able to literally do what adults do without thinking'".

But instead of supporting River's suggestion, as he always did before this point, Akein proposed to design some sacrificial concepts instead of talking with users.

> Hearing that, a high-pitch voice leaked from River before Akein finished his sentence. "Hmm, um-hum!" (1.68 z-score). River opened himself up to go with the partner's thinking process to ideate.
>
> However, after a few minutes of exploring the solution space, River was channeled back to think of talking to users, because one of their concepts was hinged on the question "when kids transition from baby food [to normal meals]". "That", points River to the TV screen which shows a paused video of children at eating, "looks like my plate, only it's not working!... And we literally had to separate ourselves from the kids because it doesn't work… Ah, I… I would love to ask about when you eat together, when you don't".
>
> In response, Akein insisted on spending the limited time on concept generation and preparing for user-testing. River attempted to persuade the partner again: "We've thrown some ideas out around this stuff, and I'd love to see — are we even close to be on track", and he added, "our hypothesis is that they [parents and kids] are struggling eating together. I just, I want to validate that".

But Akein did not give in. “Okay” (2.14 z-score), responded River disappointedly.

In the first high−pitch reaction — “Hmm, um-hum!” (1.68 z-score), River was verbally positive, but his high vocal pitch suggests surprise and discomfort. As an experienced designer, River was able to regulate his emotions and make conscious balance between maintaining collaboration and making the right pathway. Despite a deep conviction in his own way, that it was not wise to go too deep in concept generation without validating their underlying assumptions with actual users first, River started navigating the conflict with a positive attitude and open mind. By cooperating on ideation, River built a new mental container that would accommodate product ideas while not stretching his own convictions too much. But that only led to building up more tension. At the end of the brief ideation when Akein still expressed strong interest to work on concepts, River realized they had different conceptions about “concept” — while for River they already had some high-level ideas, to his partner these were not concrete concepts ready to be tested with users.

Negative signals from Akein deflated River’s positive energy and drove up the stress again, as he tried to restore the team morale by disobeying his unwillingness to follow Akein’s direction. It also had a pivotal effect on River’s subsequent performance. He was less curious about the task in general and less engaged behaviorally. He changed from the active posture of standing up and working by the whiteboard back to sitting down most of the time. He also laughed much less and was more immune to novel situation according to the high-pitch-incident criteria, as shown Figure 6b, in the second half of the design process.

In the end, the two designers delivered a concept of a utensil-food set for young kids to eat, socialize and learn with their families. The utensils were designed to be used with specific family foods, optimized for little hands in terms of size, shape, consistency and viscosity. River regarded this result as “pretty good”. During the user testing, a parent expressed her interest to buy this product if it were on the market. But for a designer who regarded “keeping the energy up” as crucial for the success of a project, the project in River’s mind ended in “anticlimax”.

5.3.2 Comparison of situated emotion between River and Jurian

In marked contrast, Jurian had a very different way of engagement in design. In the post-project interview, Jurian revealed that, as a professional, he tended to detach himself from the work, as for him, design was all about applying “the process” and “rules of engagement”. He positioned the design task as a work responsibility he had agree to work on. Different than River who went with his flow, Jurian carefully devised the timeline, planning how much time is used to scope the problem and when to switch to prototyping. Although neither of the designers were parents themselves, compared to River who developed a relationship with the design task, Jurian admitted in retrospect that he did not like the design topic at all because he had no affinity with children. Despite this disaffection, Jurian started off the task with high yet controlled engagement. In fact, he was one of the few who took careful notes in reading the design materials.

Semantic analysis also offers a perspective to understand Jurian's controlled way of engagement through being professional: he consistently used positive words as he dealt with novel situations, such as "wow this is great!" and "interesting!". This contrasts with River's use of diverse word expressions as analyzed earlier.

The contrasting design intentions of the two designers allow us to make sense of the contrasting levels of SCR frequency. Jurian's SCR frequency maintained well-controlled and low with only a few exceptions. Notably, Jurian had a stressful experience with an unexpected user-testing situation as his rules of engagement were broken. This highly emotional experience is reflected by his rising vocal pitch, frequent high-pitch talking (Figure 6a), which co-occurred with a SCR peaking (Figure 6b) during user testing detailed analysis can be found in Table 3.

We have shown how situated emotions can be identified, characterized and analyzed through contextualization and triangulation with multiple measures, despite apparent discordance. We see from River's example how a designer's emotions, thoughts and actions took place in a distributed manner through the unfolding socio-physical activities over time, and especially how each of the situated emotional experiences channeled the designer into a particular design direction out of the many possibilities embedded in the ill-defined design task.

5.4 Emotion engagement correlates with design framing

In this section, we examined the hypothesis that intense emotional engagement positively correlates with change of design frame through multilevel model analysis. The intra-class correlation (ICC), i.e., the ratio of the random intercept variance (between-person) to the total variance (between- and within-person) is 15.87%. This means the within-person variance is 84.13% with lots of within-person variance to model. All the variables are at individual-level, aggregated at the time scale of dyad-specific design phases and social situations. Because SCR frequency shows no time covariation with probability of increasing use of new words, we have used vocal pitch metrics as main independent variables, and SCR frequency as a control variable.

Table 3 Comparison of situated emotions between River and Jurian

Design stage	*River* "going along where your energy is"		*Jurian* "knowing the rules of engagement"	
	Physiological measures	*Self-report & Video*	*Physiological measures*	*Self-report & Video*
Initial exploration	**High stress** ❑ Densely populated with 40 ppm of SCR freq. or more ❑ Low average pitch, but dense high-pitch incidents of surprise and curiosity in engagement with material and work partner	Early interaction with the unfamiliar partner was regarded most stressful - "figuring out the partner was the most important piece of getting through"; Meanwhile, River started to get intrigued and energized by the design materials	**Low-to-medium stress** ❑ Spike of SCR freq. above 20 ppm	First time working with a much more senior partner on a design topic he did not like, Jurian was able to manage stress by detaching himself from the task and being "professional"
Need-finding	**High surprise and confusion** ❑ High, but lowered SCR freq. ❑ High pitch with densely populated high-pitch incidents of surprise and confusion	Frequent stimulation by user behaviors in engagement with video materials. Meanwhile, stress level dipped as River learned more about the work partner	**Momentary high energy, surprise and confusion** ❑ Spike of SCR freq. above 20 ppm ❑ Spike of vocal pitch and clustering of high-pitch incidents	Jurian had high engagement with video materials. Behaviorally, he stood up close to the TV screen and gestured a lot. He was surprised and curious to learn a few user behaviors with the work partner
Ideation and prep for user-testing	**High stress** ❑ 40 ppm of SCR freq. or more ❑ Low average pitch ❑ Decreased freq. of laughter hereafter	After the tense conflict with the partner, River was under great stress of restoring team morale by disobeying his unwillingness to design products without fully understanding user needs	**Low stress & energy** ❑ Low-to-medium SCR freq. below 5 ppm ❑ Occasional high-pitch incidents ❑ Rising pitch just before user testing	Jurian stayed low arousal; His pitch rose as he saw the users from the glass door, as he was surprised to see a baby, since the solution was designed for bigger kids
User testing	**Interest** ❑ Dipped SCR freq. ❑ High pitch with a clustering of high-pitch incidents as stimulated by interaction with users	River sharpened tone a few times to ask the parents questions, e.g., "*Do you guys ever eat together at that table or do you eat separately?*" (z-score: 1.67). The answer to this question was regarded critical to validating his assumption	**High stress, surprise and curiosity** ❑ Medium-to-high SCR freq. above 15 ppm ❑ High pitch with frequent high-pitch incidents	Jurian's rules of engagement was broken here, partly because of the unprepared user-testing situation. Jurian voice got tensely high as he talked with the two parents. He reported it was good surprise for the project

(*continued on next page*)

Table 3 (*continued*)

Design stage	*River* "going along where your energy is"		*Jurian* "knowing the rules of engagement"	
	Physiological measures	*Self-report & Video*	*Physiological measures*	*Self-report & Video*
Reflection and Pitch delivery	**Low interest** ❑ Lowest SCR frequency ❑ Lowest average vocal pitch	River believed the time after needfinding was not spent correctly. As a result, the project ended in his mind in an "anticlimax"	**Medium energy** ❑ Fluctuating SCR freq. ❑ Clustering of high-pitch incidents in pitch delivery	Jurian was curious to share his new findings from the user-testing with his partner; His voice pitched high a few times as he performed the delivery talk

Note: Quotes are from the retrospective emotional experience.

Table 4 Multilevel model over vocal pitch variability for probability of acute increase of new words

Parameter	*Model 1* *Not including between-designer difference*	*Model 2* *Including between-designer difference*	*Model 3* *Including other control variables*
	Estimate (SE)	*Estimate (SE)*	*Estimate (SE)*
Fixed effects			
Intercept	11.55***(1.71)	11.71***(1.75)	14.57***(2.58)
Time	−0.41***(0.05)	−0.40***(0.05)	−0.40***(0.06)
Pitch variability	20.10***(2.55)	19.74***(2.59)	18.81***(2.92)
Social interaction			0.57 (1.74)
SCR frequency			0.01 (0.04)
Vocal intensity			−4.91*(2.59)
Random effects			
Intercept variance	2.73** (3.12)	4.51** (9.59)	4.51** (9.07)
Pitch variance		2.56** (9.68)	6.39** (10.20)
Residual variance	52.03**(4.67)	52.04** -	53.98** -
AIC	1801.37	1809.15	1671.40
-2LL	1787.38	1787.15	1645.4
Observations	262	262	241
Designers	10	10	9

Note: *p < 0.1; **p < 0.05; ***p < 0.001. AIC = Akaike Information Criterion; −2LL = −2 Log Likelihood, relative model fit statistics. Data from designer Phi is not included in Model 3 due to EDA measure error. Probability of acute increase of new words uses Log(x+1) transformation. Pitch variability uses z-scores (standardized score within-person) after standardization between 0 and 1 across person. Social interaction is a dummy variable, where interaction with users is coded as 1 and interaction with teammate only is coded as 0. Time is at the scale of detailed design phase level. Two other vocal pitch metrics − mean vocal pitch and probability of high-pitch incident − achieve similar modeling output as vocal pitch variability.

Three multilevel models were run, as shown in Table 4. The basic version (Model 1) only considered the intercept and time as between-person variables. Model 2 also included the main independent variable in the between-person level to see how the within-person association in Model 1 differed across the 10 designers. Model 3 added other control variables to test their moderating effects. The results show that across all models, vocal pitch variability is associated with probability of acute increase of new words. More specifically, as vocal pitch increased by 1 unit (i.e., vocal pitch increases by 1 standard deviation), the likelihood of a designer's increasing use of new words rose by 18.43%, after adjusting for variable transformation.

The likelihood of increasing new word usage also has a weak negative association with time, as well as vocal intensity, such that the quieter a designer spoke, the higher chance of new words they used. Neither stress/energy level, as indicated by SCR frequency, nor social interaction proves to be relevant. The relationship differs across the 10 designers as shown in Model 2. Statistically, the random effect (i.e., between-person difference) is $\sigma = 1.6$, and there is a very large confidence interval (95% CI = [0.04, 64.54]), which is partly

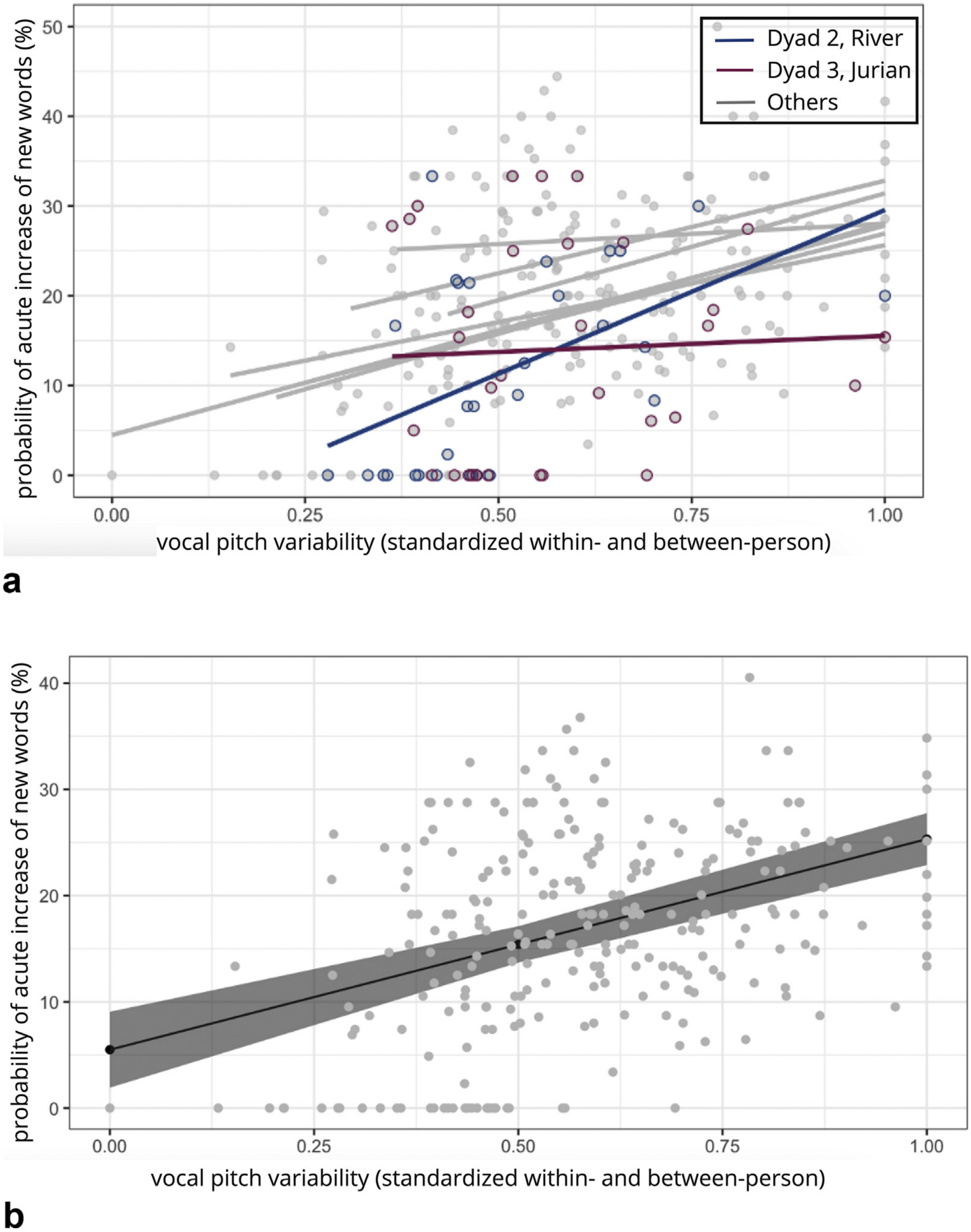

Figure 7 a. Between-designer differences of the relation between vocal variability and probability of acute increase of new words. The linear regression predictions for River and Jurian are highlighted. Positive slopes are observed in the linear regressions for all 10 designers, suggesting a likely convergent behavior of vocal pitch in relation to word usages across designers. The person-by-person analysis is nonsignificant in the cases of designers in dyad 1 and Jurian, which could be due to lack of data in each designer and/or individual difference. b. Prediction with error bar based on model 2. The result suggests the designer's intense emotional engagement, as suggested by vocal pitch variability, positively correlates with change of design frame, as characterized by emergence of new words that haven't appeared before this time

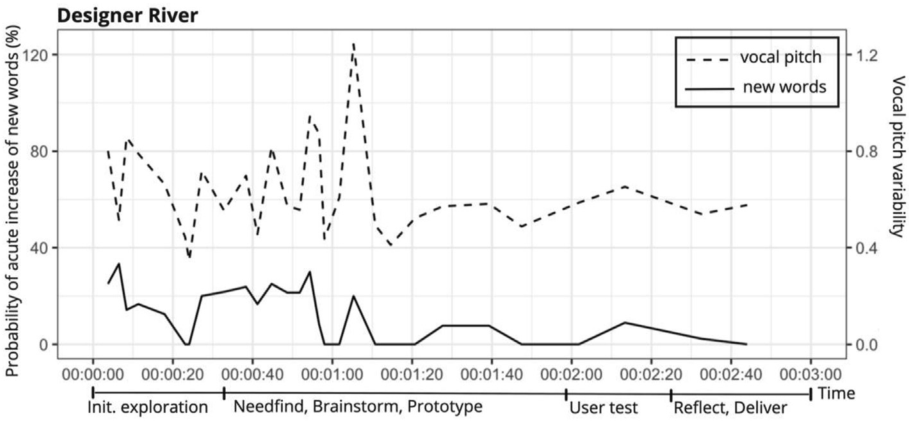

Figure 8 Time covariation between vocal pitch variability and probability of increasing use of new words, aggregated at the time scale of design phase, from designer River. Pitch variability is standardized within-person

caused by the very small number of designers included in the multilevel model. To exclude the possibility that the relationship shows a different pattern when analyzed person by person, we estimated the relationship using linear regression for each designer separately, and the result (Figure 7a) suggests a likely convergent behavior of vocal pitch in relation to word usage across designers in our current sample.

Figure 8 shows how vocal pitch variability covaries with probability of increasing use of new words along time. Here, River had higher probabilities of using new words during the needfinding phase, which co-occurred with spikes of vocal pitch variability. River's earlier example of "kidification" shows exactly how intense emotional engagement (e.g., increase of pitch variability) was associated to change of design frame, as indicated by emergence of new words around bottles, comparison between adult products and baby products and new concepts around kidification.

Together, the multilevel model result shows that a designer's intense emotional engagement, as suggested by vocal pitch, relates to change of design frame, as characterized by emergence of new words that haven't appeared before this time.

6 Discussion

Historically, design researchers have had more access to, and have placed more emphasis on, cognition and behavior in design processes. This bias is shared across disciplines. Lev Vygotsky (1962, cited in Roth & Lee, 2007) wrote more than five decades ago:

> We have in mind the relation between intellect and affect. Their separation as subjects of study is a major weakness of traditional psychology since it makes the thought process appear as an autonomous flow of 'thoughts thinking themselves,' segregated from the fullness of life, from the personal needs and interests, the inclinations and impulses, of the thinker (p. 8).

Our research addresses this bias by validating what Dong et al. (2009) have argued, that emotion is constitutive of design thinking. We have proposed a systematical mixed-method approach to study situated emotion. The novel approach is validated in a study of experienced designers in the context of collaborative design. We have also presented qualitative evidence of the interrelation between emotional engagement and the evolution of a design. This finding is further supported by statistical results from a different analytical perspective.

6.1 The study of situated emotion

The under-defined construct of emotion is composed of complex social, cultural, psychological and biological elements. So, *what is it that is measured in relation to emotion?* The current measures each sense a different aspect of emotion.

Retrospective emotional experience is subjective, event-contingent, low-granularity and lacks sharp time boundaries. Without careful nudging from the researcher, participants tend to focus on a high-level set of recollections, which does not directly translate into the level of detail of objective data. Yet, the advantage of self-report, as well as audio-visual data, is that they give researchers special access to symbolic values and sociocultural complexities associated with emotion to make sense of the series of lifeless numbers that represent physiological measures.

The physiological measures are objective, high-grained, context-independent measures of emotional arousal. Previous studies show EDA is a response system sensitive to a wide range of stimuli, making it difficult to draw any conclusive interpretation about its source of stimuli (Dawson et al., 2017). When interpreted under the current context, we find the level of SCR frequency matches well with subjective perceptions of stress and energy, the causes of which could be partly traced by retrospective self-report. Vocal pitch, on the other hand, is considered in literature as "less controllable and more leaky channels" of high emotions that are not revealed by verbal content or facial expressions associated with the message (Zuckerman & Driver, 1985; Dietrich et al., 2019). This is consistent with our finding that vocal pitch reliably points to high-emotion engagements that are elicited by unexpected stimuli. Vocal pitch characterizes engagement with design from a slightly different perspective than SCR frequency, and the general discordance is in part because novel situations that typically trigger high pitch may not cause or co-occur with energy/stress, depending on person and context. For instance, in participants that generally had high levels of energy/stress, momentary highly emotional engagements were not easily distinguishable. We also used a semantic approach to diagnose emotional arousal and valence expressed in words, which is found to reveal a designer's idiosyncratic verbal behaviors.

How are these measures relevant to designers' situated emotion? There is growing evidence that emotions are psychological constructions in situation (Barrett, 2017; Russell, 2003; Mesquita & Kawasaka, 2002; Mesquita & Boiger, 2014). In this view, designers construct emotions from sociocultural cues and their own bodily feedback. In the process of doing so, they amplify certain emotions that are valued in their larger sociocultural context and the immediate socio-physical context. Employing this situated view, we have applied each of the emotion measures to simplify the many vectors of emotion and its entanglement with the context. Some of the measures sense emotion with a number (e.g., vocal pitch), and some extract the constructed emotional experiences by retrospective reconstruction (e.g., post-project interview). A good understanding of situated emotion and its role in design is dependent on creating an accurate and relevant account with the data coming from different perspectives.

6.2 Situated emotion, design, and design ability

How does emotion relate to design and design ability? Based on the current study, we can infer that emotion is part of a designer's design ability. First, as the designer accumulates experience and expertise, their emotional tendency is also formed as part of the process and could act as a facilitating or restraining force in their design work. Second, the designer's moment-by-moment emotion is shaped in ways that are consistent with his/her ability; and it also functions in ways that are consistent with his/her ability. We elaborate the two points below.

In the distinctive case of River and Jurian, the two designers formed very different emotional tendencies. River has learned to value his idiosyncratic emotional engagement and understand the learned benefits of emotional energy to the progress of design. Jurian, on the other hand, has learned to value the efficiency and effectiveness brought by detaching personal emotions in doing design. Both designers have shaped, through accumulated experience, the trusted ways of emotional engagement that would best facilitate their interaction with the socio-physical context and design process. It would also act as a restraining force as exemplified in River's case when his situated emotional energy was deflated.

At the micro-level, the emotions that are situated in specific socio-physical contexts would also reveal the designer's ability. This is made evident, for instance, from River getting curiously stuck on a seemingly trivial observation about a water bottle for children. River's ability was in navigating the emotional tension between maintaining healthy teamwork and following the right design pathway. Examples can also be found in other designers. For Jurian, the stress during the user testing did not drive him to look for ways to quickly get out it (what a novice designer would typically do), rather, they signaled for him useful feedback for reframing and improving their design concepts. The emotion was constructed and experienced as positive discomfort to be embraced rather than a resignation. In another highly emotional case, Narra had a strong emotional reaction to a parenting issue that eluded the eyes of most other designers. These examples show how the experienced designers' moment-by-moment emotional reactions are responsive to their experienced views of design. This is further supported by the statistical modeling result: in resonance with high-emotion moments of engagement, which we understand as surprise, confusion, and curiosity, is a high likelihood of the designer's frame of mind undergoing changes.

We have used multilevel modeling to analyze the intra-individual relationships between emotion and design. This linear-regression tool is relatively simple to explain the potentially complex phenomenon. The shape of the raw data (Figure 6b) suggests a possible quadratic curve instead of a linear line, so that change of design frame depends on an optimal emotional engagement (i.e., moderately stimulated) rather than the most intense emotional engagement. In any case, designers' emotion has been shown to change in meaningful ways together with their knowledge-in-action.

Many scholars have argued that, in John Dewey's words, "thinking starts with felt difficulty" (Dewey, 1910). Others have argued that emotion assists cognition and sometimes mediates effective action without cognition (Damasio, 2006). These perspectives do signify the importance of emotion, and they might be convenient to explain how "gut feelings" would lead experienced designers' to elegantly solve tricky problems. However, the discussion of which triggers which between emotion and cognition is found less relevant from the perspective of situated emotion (see more discussion in Hoemann & Barrett, 2019). The situated view we are taking, together with evidence from the current study, supports our theory that designer emotion is constitutive of design ability, and it plays a constructive role in the production of design outcomes.

6.3 Reflection on the methodological approach

We have applied objective measures and analytical tools from affective science. In affective science, however, scholars tend to avoid applying different measures due to poor concordance (Barrett, 2017). The inconsistent result of emotion concordance in our study is consistent with findings in previous affective science research. However, because of choosing a single measure, construct validity of arousal or stress is often neglected in affective science. We argue that the validity challenge should not become the reason for not pursuing multimodal methods, because it can be adequately addressed by a complexity-embracing methodological approach. Our analysis follows the theoretical stance that emotion is situated, multidimensional and inseparable from its context. It allows the study of designers' emotions that would otherwise be unseen, undervalued or misunderstood. The current method is robust to measures that may be sensitive to real-world noises, such as EDA and vocal pitch, and is thus suitable for research where ecological validity has a non-negotiable priority.

6.4 *Reflection on ecological validity*

Traditionally, lab studies prioritize measurable, standardizable and repeatable conditions over ecological validity. While lab studies make possible a wide range of psychological measures and special instrumentations, from computer-based assessment to fMRI-based neurological measures, the rigid codification techniques employed in lab studies may greatly distort the phenomenon researchers intend to study (Crilly, 2019). Individuals are to a large extent detached from the actual emotional stimulation the real-world context would invoke. Field studies, at the opposite end of the ecological validity continuum, has the potential to address the issue. Partly due to the impracticality of many lab-based measures (e.g., facial behavior-based analysis) as well as potential problems caused by subjectivity and lack of standardization, field studies are still an underrepresented method in design research (Lawson, 2004; Crilly, 2019, p84 − p86).

The current research has sought to apply the advantage of lab studies in more ecological settings of design to represent real-world design and reveal the mundane realities of design. An ideal study would have the designers situated in real-world projects. While the current measures may be less obtrusive to design work, they are often considered intrusive to corporate confidentiality. The difficulty of video-recording design work in the real world due to confidentiality and other inconveniences has been overcome elsewhere (e.g., Christensen, Ball, & Halskov, 2017). Although we were not able to recreate a real-project experience, in reflective interviews, a few participants voluntarily expressed how real the experience felt. In particular, the involvement of real users gives the participants real-life pressure to work well and deliver well. In the meantime, because of the unconventional "lab" environment, the idea of participating in a research study and wearing emotional arousal devices dissolved gradually. With that said, naturalistic generalizability should be done with caution due to the use of technological tools and the artificial setup. To clarify, it is not our intention to generalize the designers' specific emotional patterns (such as in Table 3) to other design work settings. A designer's emotions would be jointly shaped by the socio-physical situations, their experiences, and expectations (e.g., emotional tendency). The goal was to study how to systematically examine situated emotion and understand its interrelation to design and design ability.

6.5 Implications for design practice

We hope our research can translate into a practical tool to raise emotion awareness and enhance design practice. Today, the traditional view of technical rationality (Schön, 1983; Harris, 1983) still dominates some worlds of design, such as in engineering, where emotion is often a neglected dimension. The conventional wisdom that emotion gets in the way of rational and analytical thinking is still unchallenged in some quarters. The systemic bias has to be overcome, and emotion should no longer be treated as a noisy byproduct of a designer's social—technical activity (Martin, Knopoff, & Beckman, 1998). Our research provides a unique and important lens for reimagining design practice as emotional in nature.

In addition, the current study lends insight into how to leverage emotional engagements for design reframing. This might especially be valuable for those with technical-rationality orientation and inflexible expertise schemata (Schön, 1983; Crocker, Fiske, & Taylor, 1984; Dane, 2010). The link between emotional engagement and the evolution of a design provides new insights into how to devise effective coaching interventions in design practice and education. For instance, it has implications for how to raise self-awareness of emotion and how to use emotion for design framing and reframing.

6.6 Limitations and future direction

The current work shares limitations common to studies that are based on a distinctive sample. The participants within the study may not be representative of experienced designers in general or in particular fields of design. Multilevel modeling is also constrained by the small number of participants (i.e., 10 designers), such that between-dyad analysis and between-designer analysis were problematic in the statistical sense.

Measuring emotional responses in naturalistic settings can be made possible by a few technological advances. We have adopted speech acoustics and electrodermal responses which are relatively less obtrusive and practical in field studies. Other desirable methods, such as facial expression-based emotion analysis, would require consistent recording of a single participant's face at a perpendicular angle, which is difficult to realize in a real-world situation. On the other hand, real-world noises indeed made speech data difficult to clean up and required cautious interpretation of electrodermal responses. Talking time, turn-taking and other design conversation features that are indicative of emotion were not incorporated into the speech-based analysis of emotion.

Looking ahead, ecological validity will remain a challenge in emotion research. As Lahlou (2018) has described, there is not "a single, coherent and non-contradictory account of what happened", and that "real action is often ambiguous and may have multiple determinations". Emotion can sometimes have mixed and even contradicting components simultaneously (Barrett, 2017). This is reflected in the current study as well. How to capture the nuances and contradictions of emotional experiences could be an important methodological direction. The fast-growing technological and algorithmic development will be playing a key role in research of emotion.

In design research, there are many opportunities for emotion research to tap into. In the statistical analysis of our current study, we did not directly test the effect of design expertise on the relationship between emotion and design evolution. Research in design expertise will benefit from an experiment that compares novices with experts in terms of the relationship between emotion and design behavior, process and outcome. Other important directions include, but not are limited to, furthering the understanding of how emotion plays a role in the work of expert designers, studying how emotion guides or interferes with designers' emergent design action, and what emotional behaviors are (or are not) conducive to desired project or learning outcomes. These future research efforts will be crucial to advancing the research of design.

7 Conclusion

We have presented a novel mixed-method approach to study designers' situated emotion in design. Situated emotion has been identified, characterized and understood by triangulating speech acoustics, electrodermal activity and semantic analysis with contextual inquiry of video data and retrospective self-report. We have shown how emotion in the process of design channels the formation of design. Multilevel model analysis suggests a designer's intense emotional engagement, as assessed by vocal pitch, relates to change and adjustment of design frame, as characterized by the emergence of new words not previously spoken. Our research highlights the importance of studying emotion in design research and also challenges how emotion could be studied. This will hopefully facilitate a fruitful conversation between design researchers and emotion researchers outside our field. We call for more emotion research to augment our understanding of design ability and behavior.

Note

1. https://www.axon.com/products/axon-flex-2.

2. https://www.empatica.com/en-int/research/e4/.
3. What does EDA measure? Electrodermal activity (EDA, previously known as galvanic skin response) refers to the variation of the electrical conductance of the skin in response to sweat secretion, which is produced by the sympathetic nervous system. Our skin becomes a better conductor of electricity when we receive external or internal stimuli that are physiologically arousing. For more, refer to Boucsein (2012) and others.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Ackermann, E. (2001). Piaget's constructivism, Papert's constructionism: What's the difference. *Future of learning group publication, 5*(3).

Aguera, P. E., Jerbi, K., Caclin, A., & Bertrand, O. (2011). *ELAN: A software package for analysis and visualization of MEG, EEG, and LFP signals*. Computational intelligence and neuroscience, 2011.

Ahmed, S., & Wallace, K. M. (2004). Understanding the knowledge needs of novice designers in the aerospace industry. *Design Studies, 25*(2), 155–173.

Amabile, T. M., Barsade, S. G., Mueller, J. S., & Staw, B. M. (2005). Affect and creativity at work. *Administrative Science Quarterly, 50*(3), 367–403.

Atman, C. J., Chimka, J. R., Bursic, K. M., & Nachtmann, H. L. (1999). A comparison of freshman and senior engineering design processes. *Design Studies, 20*(2), 131–152.

Atman, C. J., Sheppard, S. D., Turns, J., Adams, R. S., Fleming, L. N., Stevens, R., et al. (2010). Enabling engineering student success. In *The final report for the center for the advancement of engineering education*. Morgan & Claypool Publishers.

Balters, S., & Steinert, M. (2017). Capturing emotion reactivity through physiology measurement as a foundation for affective engineering in engineering design science and engineering practices. *Journal of Intelligent Manufacturing, 28*(7), 1585–1607.

Bamberger, J., & Schön, D. A. (1983). Learning as reflective conversation with materials: Notes from work in progress. *Art Education, 36*(2), 68–73.

Bamberger, J. (2014). The laboratory for making things: Developing multiple representations of knowledge. In *Science Teachers' Use of Visual Representations* (pp. 291–311). Cham: Springer.

Barrett, L. F. (1997). The relationships among momentary emotion experiences, personality descriptions, and retrospective ratings of emotion. *Personality and Social Psychology Bulletin, 23*(10), 1100–1110.

Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annu. Rev. Psychol., 58*, 373–403.

Barrett, L. F. (2017a). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1–23.

Barrett, L. F. (2017b). *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt.

Barsade, S. G. (2002). The ripple effect: Emotional contagion and its influence on group behavior. *Administrative Science Quarterly, 47*(4), 644–675.
Bartel, C. A., & Saavedra, R. (2000). The collective construction of work group moods. *Administrative Science Quarterly, 45*(2), 197–231.
Behoora, I., & Tucker, C. S. (2015). Machine learning classification of design team members' body language patterns for real time emotional state detection. *Design Studies, 39*, 100–127.
Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods, 190*(1), 80–91.Chicago.
Bennett, C. L., Peil, B., & Rosner, D. K. (2019). Biographical prototypes: Reimagining recognition and disability in design. In *Proceedings of the 2019 on designing interactive systems conference* (pp. 35–47).
Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc.".
Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics, 3*(2), 220–238.
Boersma, P. (2006). *Praat: Doing phonetics by computer*. http://www.praat.org/.
Bolger, N., & Laurenceau, J. P. (2013). Intensive longitudinal methods: An introduction to diary and experience sampling research. Guilford Press.
Boucsein, W. (2012). *Electrodermal activity*. Springer Science & Business Media.
Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activation. *Psychophysiology, 45*(4), 602–607.
Bucciarelli, L. L. (1994). *Designing engineers*. Cambridge, MA: The MIT Press.
Cacioppo, J. T., Berntson, G. G., Larsen, J. T., Poehlmann, K. M., & Ito, T. A. (2000). The psychophysiology of emotion. *Handbook of Emotions, 2*, 173–191.
Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1*(1), 18–37.
Cannon, D. M. (2018). *Prediction of design team performance through analysis of language use in meetings*. Stanford University.
Caruelle, D., Gustafsson, A., Shams, P., & Lervik-Olsen, L. (2019). The use of electrodermal activity (EDA) measurement to understand consumer emotions–A literature review and a call for action. *Journal of Business Research, 104*, 146–160.
Casakin, H., & Goldschmidt, G. (1999). Expertise and the use of visual analogy: Implications for design education. *Design Studies, 20*(2), 153–175.
Chai, C., Cen, F., Ruan, W., Yang, C., & Li, H. (2015). Behavioral analysis of analogical reasoning in design: Differences among designers with different expertise levels. *Design Studies, 36*, 3–30.
Chiu, M. L. (2003). Design moves in situated design with case-based reasoning. *Design Studies, 24*(1), 1–25.
Christensen, B. T., Ball, L. J., & Halskov, K. (2017). *Analysing design thinking: Studies of cross-cultural co-creation*. CRC Press.
Clancey, W. J. (2016). Creative engineering: Promoting innovation by thinking differently. In John E. Arnold (Ed.), *With an introduction and biographical essay by William J. Clancey*.
Coyne, R. (2005). Wicked problems revisited. *Design Studies, 26*(1), 5–17.
Crilly, N. (2019). Methodological diversity and theoretical integration: Research in design fixation as an example of fixation in research design? *Design Studies, 65*, 78–106. https://doi.org/10.1016/j.destud.2019.10.006.

Crocker, J., Fiske, S. T., & Taylor, S. E. (1984). Schematic bases of belief change. In *Attitudinal judgment* (pp. 197−226). New York, NY: Springer.
Cross, N. (2004). Expertise in design: An overview. *Design Studies, 25*(5), 427−441.
Cross, N., & Cross, A. C. (1998). Expert designers. In *Designers* (pp. 71−84). London: Springer.
Cross, N., Dorst, K., & Christiaans, H. (Eds.). (1996). *Analysing design activity*. Wiley.
Csikszentmihalyi, M. (2013). *Flow: The psychology of happiness*. Random House.
Csikszentmihalyi, M., & Larson, R. (2014). Validity and reliability of the experience-sampling method. In *Flow and the foundations of positive psychology* (pp. 35−54). Dordrecht: Springer.
D'Mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys (CSUR), 47*(3), 1−36.
Damasio, A. R. (2006). *Descartes' error*. London, United Kingdom: Vintage Publishing.
Davis, M. A. (2009). Understanding the relationship between mood and creativity: A meta-analysis. *Organizational Behavior and Human Decision Processes, 108*(1), 25−38.
Dane, E. (2010). Reconsidering the trade-off between expertise and flexibility: A cognitive entrenchment perspective. *Academy of Management Review, 35*(4), 579−603.
Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The electrodermal system. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Cambridge handbooks in psychology. Handbook of psychophysiology* (pp. 217−243). Cambridge University Press.
Degottex, G., Kane, J., Drugman, T., Raitio, T., & Scherer, S. (2014). COVAREP—a collaborative voice analysis repository for speech technologies. In 2014 *ieee international conference on acoustics, speech and signal processing (icassp)* (pp. 960−964). IEEE.
Dewey, J. (1910). *How we think*. Boston. MA: DC Heath.
Dietrich, B. J., Enos, R. D., & Sen, M. (2019). Emotional arousal predicts voting on the US supreme court. *Political Analysis, 27*(2), 237−243.
Dong, A., Kleinsmann, M., & Valkenburg, R. (2009). Affect-in-cognition through the language of appraisals. *Design Studies, 30*(2), 138−153.
Dorst, K. (2011). The core of 'design thinking' and its application. *Design Studies, 32*(6), 521−532.
Dorst, K. (2015). *Frame innovation: Create new thinking by design*. MIT press.
Dorst, K., & Cross, N. (2001). Creativity in the design process: Co-evolution of problem−solution. *Design Studies, 22*(5), 425−437.
Dym, C. L., & Little, P. (1999). *Engineering design: A project-based introduction*. John Wiley and sons.
Ewald, B., Menning, A., Nicolai, C., & Weinberg, U. (2019). Emotions along the design thinking process. In *Design thinking research* (pp. 41−60). Cham: Springer.
Ge, X., & Leifer, L. (2020). When tough times make tough designers: How perplexing experiences shape engineers' knowledge and identity. *International Journal of Engineering Education, 36*(2), 650−663.
Gerber, E., & Carroll, M. (2012). The psychological experience of prototyping. *Design Studies, 33*(1), 64−84.
Gero, J. S., & Milovanovic, J. (2020). A framework for studying design thinking through measuring designers' minds, bodies and brains. *Design Science, 6*. https://doi.org/10.1017/dsj.2020.15.

Gino, F., & Ariely, D. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality and Social Psychology, 102*(3), 445.

Grimm, K. J., Ram, N., & Estabrook, R. (2016). *Growth modeling: Structural equation and multilevel modeling approaches*. Guilford Publications.

Hariharan, B. (2011). *Innovating capability for (Deweyan) continuity of inquiry in the face of (Zimbardoean) discontinuity within the context of engineering education research: Fostering collaborations with underserved communities in the developing regions of the world*. Department of Mechanical Engineering Stanford University.

Harris, A. (1983). The intellectual standing of engineering design. *Design Studies, 4*(3), 147–150.

Hedman, E. B. (2014). *Thick psychophysiology for empathic design*. Doctoral dissertation, Massachusetts Institute of Technology.

Heylighen, A., & Nijs, G. (2014). Designing in the absence of sight: Design cognition re-articulated. *Design Studies, 35*(2), 113–132.

Hoemann, K., & Barrett, L. F. (2019). Concepts dissolve artificial boundaries in the study of emotion and cognition, uniting body, brain, and mind. *Cognition & Emotion, 33*(1), 67–76.

Hu, Q., Bezawada, S., Gray, A., Tucker, C., & Brick, T. (2016). Exploring the link between task complexity and students' affective states during engineering laboratory activities. In *ASME 2016 international design engineering technical conferences and computers and information in engineering conference*. American Society of Mechanical Engineers Digital Collection.

Hutchins, E. (1995). Cognition in the wild *(No. 1995)*. MIT press.

Hutchinson, A. (2018). *Designers, emotions, and ideas: How graphic designers understand their emotional experiences around ideation*. A Dissertation at Wayne State University.

Jahng, S., Wood, P. K., & Trull, T. J. (2008). Analysis of affective instability in ecological momentary assessment: Indices using successive difference and group comparison via multilevel modeling. *Psychological Methods, 13*(4), 354.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24*(4), 602–611.

Jung, M. F. (2016). Coupling interactions and performance: Predicting team performance from thin slices of conflict. *ACM Transactions on Computer-Human Interaction (TOCHI), 23*(3), 1–32.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count. *American Journal of Psychology* 263–286.

Kessous, L., Castellano, G., & Caridakis, G. (2010). Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces, 3*(1–2), 33–48.

Kilker, J. (1999). Conflict on collaborative design teams: Understanding the role of social identities. *IEEE Technology and Society Magazine, 18*(3), 12–21.

Kirsh, D. (2008). Problem solving and situated cognition. In P. Robbins, & M. Aydede (Eds.), The cambridge Handbook of situated cognition *(cambridge handbooks in psychology* (pp. 264–306). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511816826.015.

Lahlou, S. (2011). How can we capture the subject's perspective? An evidence-based approach for the social scientist. *Social science information, 50*(3–4), 607–655.

Lahlou, S. (2018). *Installation theory: The societal construction and regulation of behaviour*. Cambridge University Press.

Lang, P. J. (1988). What are the data of emotion?. In *Cognitive perspectives on emotion and motivation* (pp. 173−191) Dordrecht: Springer.
Lawson, B. (2004). Schemata, gambits and precedent: Some factors in design expertise. *Design Studies, 25*(5), 443−457.
Lewin, K. (1999). The complete social scientist: A Kurt Lewin reader. American Psychological Association.
Lougheed, J. P., Brinberg, M., Ram, N., & Hollenstein, T. (2020). Emotion socialization as a dynamic process across emotion contexts. *Developmental Psychology, 56*(3), 553.
Martinez, A. M. (2017). Visual perception of facial expressions of emotion. *Current Opinion in Psychology, 17*, 27−33.
Martin, J., Knopoff, K., & Beckman, C. (1998). An alternative to bureaucratic impersonality and emotional labor: Bounded emotionality at the Body Shop. *Administrative Science Quarterly* 429−469.
Mauss, I. B., & Robinson, M. D. (2009). Measures of emotion: A review. *Cognition & Emotion, 23*(2), 209−237.
Mauss, I., Wilhelm, F., & Gross, J. (2004). Is there less to social anxiety than meets the eye? Emotion experience, expression, and bodily responding. *Cognition & Emotion, 18*(5), 631−642.
Mesquita, B., & Boiger, M. (2014). Emotions in context: A sociodynamic model of emotions. *Emotion Review, 6*(4), 298−302.
Mesquita, B., & Kawasaka, M. (2002). Different emotional lives. *Cognition & Emotion, 16*, 127−141.
Micheli, P., Wilner, S. J., Bhatti, S. H., Mura, M., & Beverland, M. B. (2019). Doing design thinking: Conceptual review, synthesis, and research agenda. *Journal of Product Innovation Management, 36*(2), 124−148.
Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word−emotion association lexicon. *Computational Intelligence, 29*(3), 436−465.
Morozov, A., Kilgore, D., Yasuhara, K., & Atman, C. (2008). Same courses, different outcomes? Variations in confidence, experience, and preparations in engineering design. In *Proceedings of the American society for engineering education annual conference* (pp. 2008−2768).
Op't Eynde, P., & Turner, J. E. (2006). Focusing on the complexity of emotion issues in academic learning: A dynamical component systems approach. *Educational Psychology Review, 18*(4), 361−376.
Paletz, S. B., Schunn, C. D., & Kim, K. H. (2011). Intragroup conflict under the microscope: Micro-conflicts in naturalistic team discussions. *Negotiation and Conflict Management Research, 4*(4), 314−351.
Paredes, P. E., Ordonez, F., Ju, W., & Landay, J. A. (2018). Fast & furious: Detecting stress with a car steering wheel. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1−12).
Parkinson, B. (2012). Piecing together emotion: Sites and time-scales for social construction. *Emotion Review, 4*(3), 291−298.
Pea, Roy D. (1993). Practices of distributed intelligence and designs for education. In Gavriel Salomon (Ed.), *Distributed cognitions Psychological and educational considerations* (pp. 47−87). Cambridge university press.
Pekrun, R., Hall, N. C., Goetz, T., & Perry, R. P. (2014). Boredom and academic achievement: Testing a model of reciprocal causation. *Journal of Educational Psychology, 106*(3), 696.
Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion, 37*, 98−125.

Pychyl, T. A., Lee, J. M., Thibodeau, R., & Blunt, A. (2000). Five days of emotion: An experience sampling study of undergraduate student procrastination. *Journal of Social Behavior and Personality, 15*(5), 239.
Ram, N., Brinberg, M., Pincus, A. L., & Conroy, D. E. (2017). The questionable ecological validity of ecological momentary assessment: Considerations for design and analysis. *Research in Human Development, 14*(3), 253–270.
Ram, N., Gerstorf, D., Lindenberger, U., & Smith, J. (2011). Developmental change and intraindividual variability: Relating cognitive aging to cognitive plasticity, cardiovascular lability, and emotional diversity. *Psychology and Aging, 26*(2), 363.
Reisenzein, R. (2000). Exploring the strength of association between the components of emotion syndromes: The case of surprise. *Cognition & Emotion, 14*(1), 1–38.
Rieuf, V., Bouchard, C., Meyrueis, V., & Omhover, J. F. (2017). Emotional activity in early immersive design: Sketches and moodboards in virtual reality. *Design Studies, 48*, 43–75.
Roth, W. M., & Lee, Y. J. (2007). "Vygotsky's neglected legacy": Cultural-historical activity theory. *Review of Educational Research, 77*(2), 186–232.
Runco, M. A. (1999). Tension, adaptability, and creativity. *Affect, Creative Experience, and Psychological Adjustment* 165–194.
Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161.
Russell, J. A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin, 115*(1), 102–141. https://doi.org/10.1037/0033-2909.115.1.102.
Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*, 145–172.
Safin, S., Dorta, T., Pierini, D., Kinayoglu, G., & Lesage, A. (2016). Design flow 2.0, assessing experience during ideation with increased granularity: A proposed method. *Design Studies, 47*, 23–46.
Sas, C., & Zhang, C. (2010). Investigating emotions in creative design. In *Proceedings of the 1st DESIRE network conference on creativity and innovation in design* (pp. 138–149). Desire Network.
Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
Schutz, P. A., DeCuir-Gunby, J. T., & Williams-Johnson, M. R. (2016). Using multiple and mixed methods to investigate emotions in educational contexts. In *Methodological advances in research on emotion and education* (pp. 217–229). Cham: Springer.
Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge university press.
Todorova, G., Bear, J. B., & Weingart, L. R. (2014). Can conflict be energizing? A study of task conflict, positive emotions, and job satisfaction. *Journal of Applied Psychology, 99*(3), 451.
Tsai, J. L., Knutson, B., & Fung, H. H. (2006). Cultural variation in affect valuation. *Journal of Personality and Social Psychology, 90*(2), 288.
Valentine, M. A., Nembhard, I. M., & Edmondson, A. C. (2015). Measuring teamwork in health care settings: A review of survey instruments. *Medical Care, 53*(4), e16–e30.
Valkenburg, R., & Dorst, K. (1998). The reflective practice of design teams. *Design Studies, 19*(3), 249–271.

Vallet, F., Eynard, B., Millet, D., Mahut, S. G., Tyl, B., & Bertoluci, G. (2013). Using eco-design tools: An overview of experts' practices. *Design Studies, 34*(3), 345–377.

Villanueva, I., Campbell, B. D., Raikes, A. C., Jones, S. H., & Putney, L. G. (2018). A multimodal exploration of engineering students' emotions and electrodermal activity in design activities. *Journal of Engineering Education, 107*(3), 414–441.

Vincenti, W. G. (1990). *What engineers know and how they know it analytical studies from aeronautical History*. Baltimore: Johns Hopkins University Press.

Voigt, R., Podesva, R. J., & Jurafsky, D. (2014). Speaker movement correlates with prosodic indicators of engagement. In *Speech prosody, Vol. 7*.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151.

Xue, H., & Desmet, P. M. (2019). Researcher introspection for experience-driven design research. *Design Studies, 63*, 37–64.

Yilmaz, S., & Seifert, C. M. (2011). Creativity through design heuristics: A case study of expert product design. *Design Studies, 32*(4), 384–415.

Zhou, J., Phadnis, V., & Olechowski, A. (2019). Analysis of designer emotions in collaborative and traditional computer-aided design. In . *International design engineering technical conferences and computers and information in engineering conference, Vol. 59278*. American Society of Mechanical Engineers. V007T06A043.

Zuckerman, M., & Driver, R. E. (1985). Telling lies: Verbal and nonverbal correlates of deception. *Multichannel Integrations of Nonverbal Behavior* 129–147.