Creative Sense-Making: Quantifying Interaction Dynamics in Co-Creation

Nicholas Davis, Chih-Pin Hsiao, Kunwar Yashraj Singh, Brenda Lin, Brian Magerko
Georgia Institute of Technology
School of Interactive Computing
{ndavis35, chsiao9, kysingh, blin15, magerko}@gatech.edu

ABSTRACT
This paper describes a new technique for quantifying interaction dynamics during open-ended co-creation, such as collaborative drawing or playing pretend. We present a cognitive framework called *creative sense-making*. This framework synthesizes existing cognitive science theories and empirical investigations into open-ended improvisation to develop a method of quantifying cognitive states and types of interactions through time. We apply this framework to empirical studies of human collaboration (in the domain of pretend play) and AI-based systems (in the domain of collaborative drawing) to establish its validity through cross-domain application and inter-rater reliability within each domain. The creative sense-making framework described includes a qualitative coding technique, interaction coding software, and the cognitive theory behind their application.

Author Keywords
Interaction; Interaction Dynamics; Creativity; Collaboration; Co-Creation; Sense-making

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous;

INTRODUCTION
Collaboration is a powerful way to inspire and support creativity. During creative improvisational collaboration, a new form of distributed creativity arises that can lead to emergent, dynamic, and unexpected meaning to support creativity in new ways [29]. Subsequently, the field of computational creativity is beginning to explore how intelligent agents might collaborate with humans during their creative process in co-creative systems, i.e. systems that contribute content to a shared creative product with the user. A cognitive theory describing collaborative creativity would help design and evaluate co-creative systems. In particular, measuring collaboration outcomes involves being able to quantify *interaction dynamics*, e.g. the rhythm of interaction, style of turn taking, and manner in which participants are mutually making sense of a situation through time [9]. However, there is a gap in the literature about quantifying and evaluating open-ended creative collaboration.

C&C '17, June 27-30, 2017, Singapore, Singapore

ACM ISBN 978-1-4503-4403-6/17/06…$15.00
http://dx.doi.org/10.1145/3059454.3059478

We have seen evidence of how collaboration leads to dynamic and emergent meaning structures that inspire novel ideas during empirical studies of pretend play [7] and collaborative drawing [8]. We have also found how collaborators often provide unexpected ideas and thus lead to surprising results. The same dynamism and flexibility that make creative collaboration so effective, also unfortunately make it especially hard to quantify and evaluate.

This paper applies the cognitive science theory of enaction and its conceptual framework, participatory sense-making, to the domain of open-ended creative collaboration and then formalizes this theory as computational models of creative collaboration. The described qualitative coding and analysis technique, called the sense-making curve analysis, provides a means to rapidly and reliably quantify interaction dynamics continuously through time. This temporal data can be mathematically analyzed using continuous functions (e.g. moving averages, integrations) to formally classify different sense-making strategies and trends in creative collaboration.

The creative sense-making framework is applied to empirical studies of human collaboration (in the domain of pretend play) and technical systems (in the domain of collaborative drawing) to establish its validity through cross-domain application and inter-rater reliability within each domain. With the creative sense-making framework, we begin to offer a set of techniques and tools to begin investigating the overarching research question: how do humans collaborate in open-ended improvisational creativity, and how can we design co-creative agents to achieve similar benefits as human collaboration? The contributions of this paper include introducing the creative sense-making theoretical framework, coding technique, and analysis approach.

RELATED WORK

Several metrics have been developed relating to creativity and technology. The creativity support index (CSI) is a psychometric survey instrument that measures the effectiveness of a creativity support tool for assisting users engaged in creative work [5]. Protocol analysis has been used as a method for evaluating and comparing how different user interfaces and input methodologies affect creative cognition [24]. Some aspects of the user's creative process can be quantified by logging user data from creativity support tools, such as measuring the emergence of new ideas and ideation strategies [23], the novelty and surprise of designs [27], the effectiveness of a creative collaboration [8], and the relative entropy of ideas generated throughout a design protocol [18]. Biometric sensors, such as EEGs, have also been employed to quantify *in the moment creativity* (ITMC) by classifying periods of heightened creativity based on physiological markers [4].

Due to the complexity and open-ended nature of creative activities, researchers generally employ a mixed-method approach of data triangulation that draws on multiple sources of data to analyze and evaluate how a technological intervention affects the creative process [4]. While these creativity research methods provide insight into an individual's creative process and tools utilized during creativity, evaluating creative collaboration presents unique challenges around understanding how collaborators coordinate in the moment to co-construct shared meaning throughout a creative collaboration.

The enactivist paradigm in cognitive science has made significant advancements in terms of understanding how meaning emerges through interaction, both by an individual agent and through social coordination [3,10]. These researchers propose a novel theoretical framework focused on the idea of sense-making, whereby an agent gradually casts a web of significance and meaning onto the world through interacting with the environment (and other agents within it) to determine meaningful regularities [33–35]. The theoretical framework of sense-making is conceptually robust, replete with a vocabulary and paradigmatic viewpoint for understanding cognition and interaction through the lens of socially emergent and dynamic meaning constructs [12,20,36]. However, there is a significant gap in the field regarding quantifying the interaction dynamics of sense-making during complex and open-ended activities, such as improvisational creative collaboration. In the enactivist literature, there are generally two approaches to quantifying interaction dynamics and sense-making: *perceptual crossing* and *traditional qualitative analysis*:

Perceptual Crossing Methodology

Perceptual crossing is a type of participatory sense-making and experimental apparatus recently introduced into the cognitive science literature to quantify interaction patterns directly within an artificially constrained virtual environment [1,10,11,17]. Researchers studying perceptual crossing utilize a minimalist virtual environment in which two agents (either both human, both artificial agents, or a combination thereof) perform a spatial participatory sense-making task whereby they try to differentiate the motion of their partner from static objects using only tactile feedback. Each player moves their avatar across their respective screens using the mouse or arrow keys, and the participants receive a vibration if the avatar has crossed paths with something in the environment (whether it is a static object or human). Participants have to interactively explore each stimulus to determine whether it was caused by a static object or another human in a process of *participatory sense-making*. In this virtual environment, all actions are restricted to constrained digital inputs and are thus easily observable and quantifiable. This approach reveals some of the underlying processes of participatory sense-making, but it is not applicable for understanding sense-making in more complex domains, such as open-ended and creative interactions.

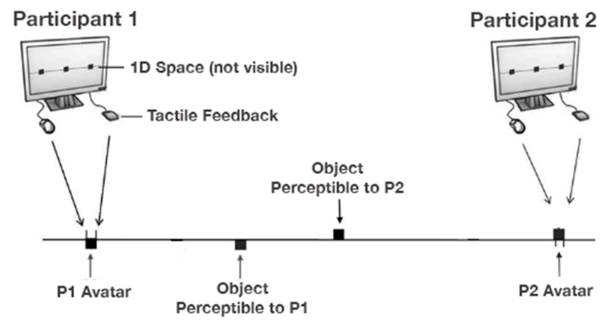

Figure 1. Perceptual crossing experimental set up (adapted from [1]). Participants interact with the virtual environment to discern static objects from other agents.

Traditional Qualitative Analysis

The second approach for understanding interaction dynamics is traditional qualitative analysis, i.e. qualitatively coding observational video data of complex social activities to interpret the types of actions and strategies recruited during participatory sense-making [22,31]. This approach is widely used in dialog and conversation analysis [32] and analyzing turn taking dynamics, such as leader/follower strategies and topic shifts throughout the interaction [2]. These investigations typically yield descriptive statistics about the types of events that occurred accompanied by a thick cognitive ethnographic description of the factors that influenced participatory sense-making, but these accounts are often difficult to quantify.

Qualitative analyses employing event-based coding practices are also common in analyzing and evaluating open-ended creativity, such as artistic and design creativity.

For example, this approach has been utilized to evaluate: collaboration practices of designers in a virtual environment by segmenting data into discrete events based on behavioral markers [26], as well as understanding open-ended artistic creativity [38,27]. Interestingly, [38] included the behavioral markers of pausing and body-repositioning (i.e. stepping back from the artwork) as important events in the coding scheme, which bears similarity to our proposed approach. The methods described in this section all yield powerful and informative descriptions about the number of times events occurred and even the order in which these events occur, but they are not specifically designed to quantify the fine-grained temporal dynamics of interaction.

CREATIVE SENSE-MAKING

Creative sense-making (CSM) bridges the two general methodological approaches described in the related work (e.g. perceptual crossing and qualitative coding) by creating a simplified qualitative coding scheme focused on sense-making that lends itself to quantification and computational analysis. The analysis demonstrates how different types of collaboration strategies and styles can be quantitatively classified through time. CSM was designed to support the evaluation of continuous interactions in co-creation, such as collaborative drawing and playing pretend. Using this method provides insight about the interaction dynamics and interaction trends in open-ended co-creation contexts.

In the CSM approach, each participant's action is categorized according to its functional role in sense-making. We employ the *free energy principle of the brain* [14–16] to develop a quantitative descriptor for these different states. The free energy theory was selected as it has neurobiological plausibility and it lends itself to quantification and continuous analysis of agents through time. The free-energy principle describes how biological systems continually strive to reduce surprise, i.e. cognition strives to create a dynamic and generative mental model that makes the environment more predictable [15,28,30]. A few critical definitions will help elucidate the theory of free energy:

Generative model: “or forward model is a probabilistic mapping from causes to observed consequences (data). It is usually specified in terms of the likelihood of getting some data given their causes (parameters of a model) and priors on the parameters” [14].

Surprise to agent: this occurs when a cognitive agent has developed a generative model of interaction and anticipates certain sequences of data using that model, but the data from the environment violates the expectation of the agent.

Free energy: “Free energy [is] an information theory measure that bounds or limits (by being greater than) the surprise on sampling some data, given a generative model” [15].

Free-energy principle: “The free-energy principle says that any self-organizing system that is at equilibrium with its environment must minimize its free energy.” [15].

Similar to sense-making, the free-energy principle claims that humans interact with the environment, through both active perception and action, to improve the accuracy of their generative model of the environment, thereby reducing *free-energy* [15,28]. When free energy is minimized, actions are generated fluidly and with ease, allowing agents to directly perceive affordances and meaning constructs in the environment that increase the order and predictability of interactions. When free energy increases, i.e. the cognitive agent becomes surprised, perception and action are utilized to help increase the accuracy, or *recognition density*, of the generative model of the environment, i.e. making the predictions of the generative model more closely match the ‘true’ conditions of the situation.

Combining the free-energy principle with the conceptual framework of sense-making enables a new method for quantifying interaction dynamics based on the relative free energy of a cognitive agent through time, as determined through behavioral markers. In the proposed approach, when agents do not have a robust generative model of the environment, they have to engage in a process of sense-making, which costs physical and mental energy. This type of sense-making can be viewed as an investment of physical and mental energy that has the potential to reduce free energy, in the long run, i.e. improving the predictive power of the cognitive agent’s generative mental model. Thus, **sense-making** is formally defined here as the process whereby a cognitive system gradually minimizes free-energy by reflecting on and experimentally interacting with the environment to build and refine a more optimal generative mental model of that environment.

Within the context of this framework, we propose a continuum of cognitive states corresponding to the functional role of sense-making in the agent’s interaction with the environment. Borrowing terminology from Glenberg [19], we refer to the ends of this spectrum as *clamped* and *unclamped* cognition. The concept of clamping represents a cognitive mechanism that helps the agent balance exploration versus exploitation to learn a better, more accurate, generative model (i.e. make sense of the environment).

Clamped Cognition: The process of maintaining or slightly refining the selected generative model assuming that it is the most accurate representation of the environment. It generally occurs after making sense of a task or activity. Behavioral markers include fluid interactions with minimal hesitation (e.g. embodied play actions, fluid drawing actions).

Unclamped Cognition: The process of changing or replacing the generative model by exploring and reflecting on the environment from different perspectives. It generally occurs during task onset and after surprises during the task. Behavioral markers include hesitation (e.g. eyes closed, confused look) and physically experimenting with the environment and viewpoint of the environment (e.g. futzing, inspecting).

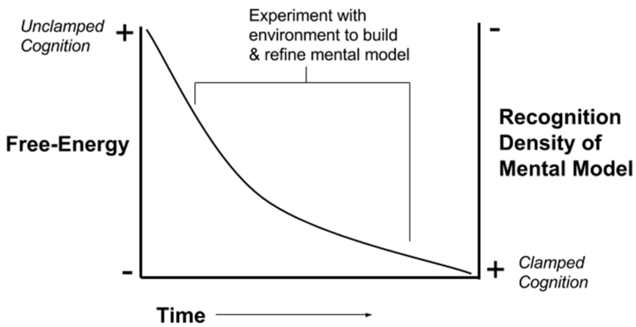

Figure 2. Visual representation showing the relationship between free-energy and clamped/unclamped cognition.

We devised a set of behavioral markers to detect clamped and unclamped cognition through a qualitative coding procedure of video data. For example, in the context of pretend play, a participant that is embodying a play character and performing fluid play actions inside the narrative world (i.e. diegetic actions) is considered clamped. Conversely, if a player is hesitating, unsure, pausing, or otherwise disengaged, this is a sign they may be actively processing and working to make sense of the situation to determine an effective strategy for moving forward, which signifies an unclamped state. Pauses and hesitations are also used in conversation analysis and analyzing interaction dynamics [2]. When players restructure or build additional meaning into the environment explicitly (i.e. extra-diegetic actions), they are actively making sense of the situation in an unclamped manner.

In creative interaction, meaning is continually shifting and evolving due to the participatory nature of improvisational collaboration [29]. However, empirical evidence [3,7] suggests that semi-stable (yet dynamic) meaning structures emerge to produce a steady state whereby both participants have a relatively robust predictive model that enables them to interact fluidly in a situated manner, without much explicit sense-making outside of fine-grained coordination. In our work, these stable units of meaning are referred to as *nucleus activities* that can grow through time as additional layers of meaning are added [7]. These semi-stationary states arise when both participants are clamped on a similar meaning structure. Conversely, participants can be confused or uncertain about what type of action to take or how to interpret their collaborator's actions due to a sparse predictive model, which would be characterized as unclamped. This type of clamping framework extends the idea behind shared meaning established in the literature about shared mental models [6,21] by taking into account the dynamic and evolutionary nature of meaning construction in open-ended creative interactions.

Within the unclamped category, there are two further distinctions that can be made to increase the granularity and explanatory power of our proposed coding technique. Unclamped actions that are meant to reduce free energy through sense-making can be either *perceptually-based* or *physical-based*. Perceptual-based sense-making relates to refining the brain's predictive model, which subsequently changes how features in the environment are perceived and interpreted. Since these processes are happening internally, they cannot be directly observed, but individuals experiencing this cognitive state display indirect behavioral markers, such as pausing, hesitating, contemplating, and looking confused, i.e. *thinking*. Physical sense-making, on the other hand, changes the structure of the environment by manipulating and modifying the environment, or by moving the body and altering what information is available to the senses, i.e. *thinking by doing*.

These two different paths of reducing free energy represent opposites that mirror the afferent and efferent flows of sensory information in a cognitive system. Perceptual-based sense-making processes change the afferent flow of predictions being generated by the brain and projected to the body (via the nervous system) by thinking and re-evaluating the situation internally. Physical sense-making processes change the efferent flow of incoming sensory information available to the agent (to generate predictions about) by taking actions that either change the environment (through manipulation) or one's vantage point of it. Thus, our proposed approach makes two chief categorical distinctions: clamped and unclamped cognition, and within the unclamped category, there is a further distinction between perceptual sense-making and physical sense-making.

Perceptual sense-making: the cognitive agent is working to internally improve recognition density of its generative mental model, i.e. *thinking*. Behavioral markers include: hesitation, eyes closed, confusion, and task disengagement in general.

Physical sense-making: the cognitive agent is exploring the environment through interaction to decrease disorder in the environment and increase the recognition density of its generative mental model, i.e. *thinking by doing.* Behavioral markers include: experimentally manipulating resources in the environment and re-positioning the body to change available sensory data.

Physical sense-making and perceptual sense-making each have a continuum of cognitive states within them ranging from clamped to unclamped cognition. While these two spectra are interrelated, and may occur simultaneously (or in very rapid succession), analysts approximate which state is dominant through time by observing mutually exclusive behavioral markers denoting what type of sense-making activities are currently observable. For example, when someone is pausing to reflect, they often briefly stop performing functional actions while they are thinking (e.g. looking away, closing eyes, etc.), which are strong indicators the person is experiencing a perceptual unclamp.

The categorical distinction between these two types of sense-making processes is reflected by assigning one type a positive value and the other a negative value both using a scale of 0 (fully clamped) to 1 (fully unclamped). The reason this positive and negative convention was adopted is because it enables quantifying interaction trends by looking at the variance of the cumulative sum of these coded values through time. From a theoretical standpoint, a value of 0 means free energy is minimized and cognition is clamped. Any deviation from 0, in the positive or negative direction, represents an unclamp event. The decision of which category initially receives a positive vs. negative sign is inconsequential if it remains consistent throughout the analysis. The important delineation here is providing granular temporal data about interaction styles and sense-making strategies through time.

The magnitude of the numerical value assigned to the unclamped action is determined by assessing the degree of unclamp based on behavioral markers. To help systematize the coding process and increase inter-rater reliability, we use discreet points on the cognitive spectrum to represent magnitude (.5 or 1 in both categories). For example, agents can pause their actions to wait for their partner to take their turn (partial perceptual unclamp), or they can be completely confused and disengaged from the play session (full perceptual unclamp). Complete confusion and disengagement should be considered more unclamped than

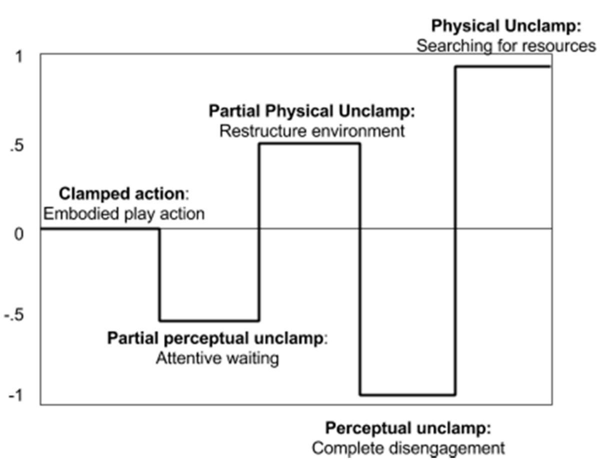

Figure 3. An example of clamped and unclamped states plotted on an axis showing how each state is coded during analysis.

attentively waiting for a partner to act. Conversely, in terms of physical sense-making, searching for new resources to introduce into the play session would be more unclamped than slightly rearranging or restructuring elements already in the play space.

When plotted along an x-axis representing time, these numerical values create what we refer to as the *sense-making curve* (as shown in Figure 3). This curve quantifies what types of actions each individual participant was engaged in throughout the course of the experiment. These curves can be mathematically analyzed and combined to quantify the interaction dynamics between the players.

To summarize, we propose a new approach to formally model the interaction dynamics of participatory sense-making. We use the qualitative video coding conventions of the sense-making curve to translate human interaction dynamics and patterns of interaction into a machine-readable format. In the next sections, we describe the web-based tool that was developed to perform this coding procedure and the analysis technique for combining sense-making curves and classifying different types participatory sense-making and styles of collaboration from sense-making curves.

SENSE-MAKING CURVE TOOL

To quantify interaction dynamics, the qualitative coding conventions described above can be applied to video analysis. To accomplish this, an analyst reviews the video and assigns a numerical value to each moment based on the type of action the participant is currently engaged in. This process is repeated for each participant in the experiment to produce unique datasets characterizing each participant's actions through time. Clamped actions are assigned a value of 0 to signify minimized free-energy, while unclamped actions are assigned a value from -1 to 1 to signify deviation from a free-energy equilibrium. The sign (positive/negative) and value of the data point corresponds to the category of unclamped action and the relative degree of the unclamp, respectively.

The prototype sense-making curve tool (shown in Figure 4) is a web-based qualitative video coding environment. To apply the qualitative codes to the video, the researcher uses the up and down arrow keys to increase or decrease the current coding value, which is visualized by a code selector slider on the right side of the screen. The system samples the value of the slider every 250ms and records that data point in an array list. Therefore, the value the code selector is currently resting at is the value that will be assigned to the time that is currently being viewed in the embedded video window.

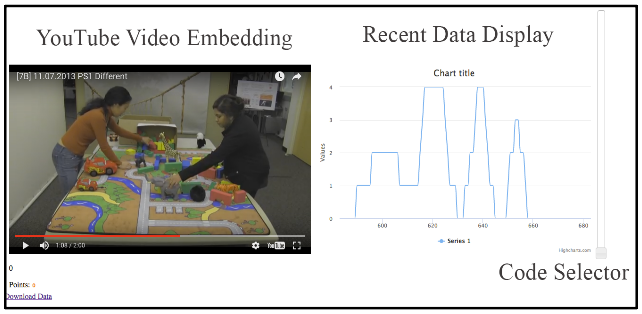

Figure 4. Prototype Sense-Making Curve Analysis Tool.

The most recent values from the array list (relative to the current point in the video that is being played) are visualized in the recent data display to the right of the embedded video. When the user scrubs the video to different time points, the data display and coding value also move to that point in the array, allowing researchers to revisit different parts of the video to re-apply codes.

The user can slow down and speed up the video playback speed. Researchers using this tool typically slow the videos down to half speed and code one participant at a time. The analyst re-watches the video to code each participant. Once an analyst was oriented to this approach, the overall code application time per video (including both participants) was approximately 4 to 1, meaning it took about 4 minutes to code every 1 minute of video data. A 4 to 1 time-efficiency ratio for qualitative video coding, and a 2 to 1 ratio for coding individual participants within an interaction, is more efficient when compared to other event-based coding methods that may take up to 30 minutes (or more) to code an individual minute of data (i.e. a 30 to 1 time-efficiency ratio).

It is important to point out that while the sense-making curve might facilitate rapid and reliable qualitative coding, the granularity of analysis between the sense-making curve and typical event-based coding techniques is much different. The sense-making curve provides continuous and targeted data about interaction dynamics, but it does not identify individual actions, i.e. the content of the actions. In future versions of this tool, we plan to allow tagging the curve with different events to combine the power of event-based coding with our proposed continuous coding technique.

The numbers representing each code on the 1 to -1 scale (according to the clamped/ unclamped convention) were each mapped to 0-4 values at the request of coders (to reduce time switching between common states in a given domain). As long as each state is consistently coded through time and the mappings to their placement on the 1 to -1 scale are stored, the numbers used in the user interface of the tool can change. This change in scale creates an interesting level of obfuscation between the analyst and the predictions of the theory and cognitive framework since the curve they are currently seeing visualized in the tool is not necessarily the final form of the curve used in the analysis (after the value transductions). It remains to be seen whether the tool is most reliable when used with the 1 to -1 scale or a 0 - 4 scale in the interface.

The data analyst's role in this context is to decide which of the five states participants are in based on behavioral markers. The design of the behavioral markers is determined by understanding what types of actions in a domain should be classified as clamped/unclamped and the degree of this association.

USING SENSE-MAKING CURVE TOOL

The coding tool and technique were used on both the pretend play dataset and the Drawing Apprentice user study data. To collect the pretend play dataset, we conduct an observational study involving 32 pairs of adult players who were asked to perform two pretend play sessions for each pair of the participants on a large play-mat with toys and stuffed animals (c.f. [7] for a full description of the study). The pretend play experimental sessions began by providing participants with a narrative prompt (e.g. monsters attack, drag racing, etc.). They were given 5 minutes for both setup and play based on their interpretation of the narrative

Clamp Value	Classifications and Behavioral Markers
1	**Physical Unclamp** **Play**: searching toy box, change locations **Drawing**: moving device, explicit feedback, verbalizations
.5	**Partial Physical Unclamp** **Play**: place items, arrange items, build new structures, extra-diegetic communication **Drawing**: aesthetic UI interactions, simulated drawing strokes
0	**Clamped Actions** **Play**: fluid and embodied play actions **Drawing**: fluid drawing actions
-.5	**Partial Perceptual Unclamp** **Play**: hesitation, holding character waiting **Drawing**: pen in hand waiting
-1	**Perceptual Unclamp** **Play**: disengaged, distracted, hands by side **Drawing**: leaning back from display, lowered pen, looking away

Table 1. Summarized coding scheme for the pretend play and Drawing Apprentice sense-making curve analysis.

Data Set	Pretend Play	Collab Drawing
Analysts	3	2
Codes Compared	19,900	362,560
Technique	Fleiss' Kappa	Kripendorff's Alpha
IRR Score	.71	.76
Interpretation	Substantial Agreement	Substantial Agreement

Table 2. Inter-rater reliability results from using creative sense-making technique to analyze pretend play and Drawing Apprentice data.

prompt. Three coders then analyzed a subset of the 64 recorded videos, resulting in a total of 128 sense-making curves produced. These three coders were able to achieve an inter-rater reliability (IRR) score of 0.71 (substantial agreement according to [13]) using Fleiss' Kappa (shown in Table 1). To calculate this score, all the coders analyzed an individual play session and the Fleiss' Kappa for the coded values of the left and right players in the session was averaged. This Fleiss' Kappa value compared the reliability of 19,990 total code applications between the analysts on the pretend play dataset.

For the Drawing Apprentice user study, six participants were asked to draw with both a human and AI drawing agent during two 12 minute drawing sessions using our prototype described in [8]. Two coders analyzed the 12 user study videos resulting in 48 total sense-making curves between the two coders with an overall inter-rater reliability of 0.76 (substantial agreement according to [13]) using Kripendorff's alpha (for two coders). For this dataset, the IRR value for coding the participant (0.64) and co-creative agent (0.86) were averaged to yield the 0.76 value for Kripendorff's alpha over the entire dataset, which consisted of 362,560 code applications (shown in Table 2). The IRR calculation for Drawing Apprentice uses more codes in its comparison because both analysts were asked to code the entire dataset since it was feasible given its size (compared to the much larger dataset for pretend play that had to be divided among the analysts).

To achieve these reliability measures, we employed the Delphi method, an iterative coding scheme refinement technique [25]. Two different teams of analysts coded the sample videos from each domain, tested their reliability, and looked for points where their coding diverged. To help visualize and quantify discrepancies between their code applications, a 2D convolution was performed between the analysts' sense-making curve data in a pairwise manner for each participant in the sampled dataset. Then, they discussed possible reasons for the divergence and added more detailed behavioral markers to the coding scheme to help ensure future reliability. After several rounds of refining the behavioral markers of the coding scheme, analysts achieved substantial agreement over thousands of code applications. In most instances in the literature, these inter-rater reliability measures are not being used on data sets with sizes this large, or with continuous temporal data. For that reason, it may be beneficial to explore additional inter-rater reliability measures to pair with the Kripendorff's alpha and Fleiss' Kappa to help evaluate inter-rater agreement through time with increased precision (such as 2D convolution and cross-correlation).

Performing the numerical transduction on the data (from the 0-4 scale to 1 to -1 scale shown in Table 1) before performing these reliability measures may also yield insights into the productivity of this method. For example, once the data is represented in the 1 to -1 convention, it can be bucketed into three categories of positive, neutral, or negative. While this reduces the overall granularity with which the data is being compared, it provides a better indicator as to whether two coders were merely off in measuring the magnitude of the unclamp versus a completely different direction.

EXEMPLAR DATASET AND ANALYSIS

This section shows the application of the sense-making curve analysis technique on a real data point from the pretend play session with 32 adult dyads. This data represents the sense-making curves coded for each of the individuals in the play session. The sense-making curves shown in Figure 5 present the qualitative coding classification of two participants during a 5-minute play session. Here, we can see both players greatly fluctuating above and below the axis (i.e. building meaning/restructure environment and waiting/disengaged, respectively) during the first third of the play session. These fluctuations may correlate to the setup period during which participants are figuring out what kind of activities to engage in during the play session by exploring the play mat, toys, and resources at their disposal. This early period represents more active sense-making where the participants are actively experimenting with the environment and directly communicating about what to do (i.e. extradiagetic communication).

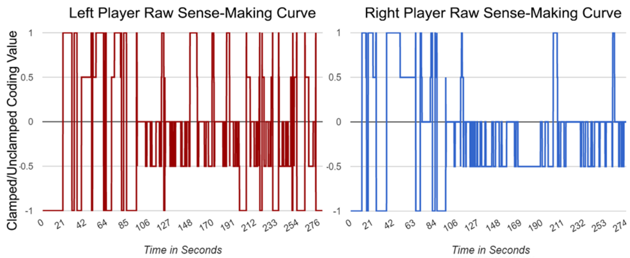

Figure 5. Sense-making curve examples from two participants during one session of the pretend play study.

After about 90 seconds (about a third of the way through the x-axis), both participants begin to engage in play activities and their curves fluctuate more closely around the 0 axis (i.e. clamped cognition), cycling between 0 (engaged in fluid play activity) and -.5 (attentively waiting). This trend indicates some degree of turn-taking is beginning to emerge at this point. It is important to note that the left player continues to deviate greatly from the 0-axis at distinct points in both the positive and negative directions, suggesting that this player was more actively involved in adding new elements into the play session as well as disengaging more often, especially in the last third of the play session.

Calculating the cumulative integral of each of these sense-making curves provides a more global picture of the participants' activities through time. For this exemplar demonstration, the MATLAB cumsum function was utilized to approximate the integral. In a more precise analysis, the MATLAB function cumtrapz was used to increase the accuracy of the integral approximation.

The cumulative integral approximations shown in Figure 6 provide an easier way to visualize how each participant was contributing to the overall flow of the play session activities. From this representation, it becomes clear that both players experienced a relatively dramatic initial hesitation and waiting period followed by a period of building meaning in the environment. However, after this initial phase, their actions seem to diverge as the right player spends more time watching or hesitating, while their partner, the left player, continues to build new meaning and engage in play (as represented by periods of rising during the cumulative integral and holding steady, respectively).

Combining these two cumulative integrals together provides information about how both participants' actions related to each other through time. To accomplish this operation, two one-dimensional arrays are added together. This combined cumulative integral can be used to identify interaction trends involving both players.

The graph in Figure 7 shows the combined cumulative integral of the participants' sense-making curves, which shows the overall activity through time with respect to both players' actions. We refer to this curve as the *creative*

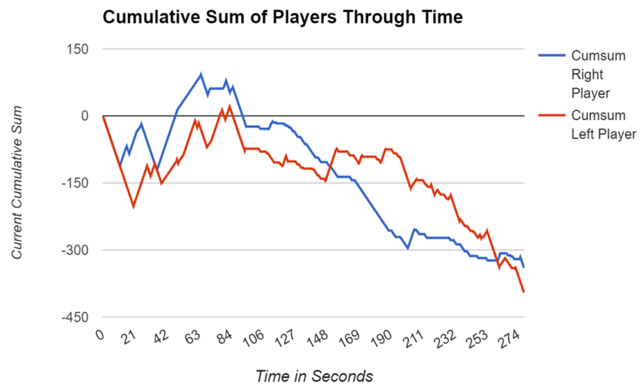

Figure 6. Running cumulative sum (integral approximation) of the right and left players' sense-making curves

trajectory because it can be used to help identify interaction trends and patterns in the overall creative flow, or trajectory, of the play session. For this exemplar demonstration, these trends are visually identified and depicted on the graph. Each of these trends theoretically corresponds to different ways of collaborating as well as helping to identify when coupled play began to arise between the players (the orange arrows in the graph). In the actual analysis, more precise computational modeling techniques are used, similar to the conventions used in stock market analysis to identify buy, sell, and hold signals from continuous stock data. The trends visualized on this graph can be summarized as follows:

Blue Line: Both players are hesitating or waiting. This classification will be most distinct when both players are completely disengaged for prolonged periods of time as they will both have values of -1 during those times. When both players are coded as -1, the cumulative integral is reduced sharply, as indicated by the large negative slope in the first three blue arrows. However, the value of the cumulative integral can also fall when both players are either attentively waiting (coded as -.5), or one is attentively waiting while the other plays (coded as 0). In

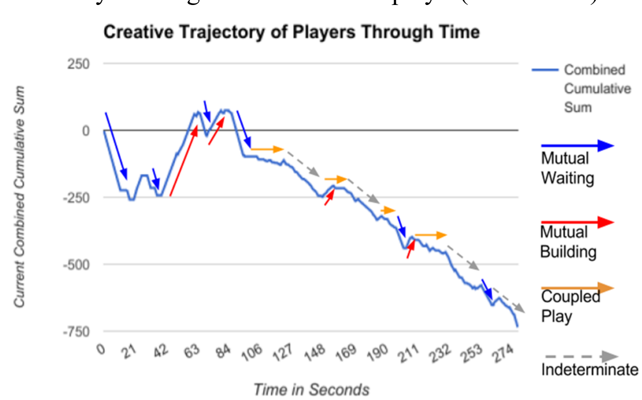

Figure 7. Sum of both players running integrals creates a *creative trajectory* depicting interaction trends between players through time

this case, the slope of the creative trajectory will not fall as fast, such as the last blue line on the right side of Figure 7.

Red Line: Both players are building meaning or gathering resources. This classification corresponds to both players actively engaging in a sense-making process (unclamped cognition) through interactively exploring their environment by building new structures in the play area or looking for new resources to add to the play area. When the slope of the creative trajectory is positive, both players are most likely engaged in unclamped actions (rated as +.5 or +1) to make the cumulative integral rise. A high positive slope would indicate that both players are actively looking for new resources (i.e. rummaging through the toy box) to add to the play mat. A smaller positive slope may indicate building or tweaking elements that are currently on the play mat.

Orange Line: Clamped quasi-stationary play state. This classification represents some type of coupled play interaction between the participants. There are many different types of coupling that can result in this type of semi-stationary state where the slope of the creative trajectory curve is near zero. To classify this state, the derivative of the creative trajectory curve can be calculated and those areas of the derivative curve that are near zero within a threshold of standard. These couplings indicate regions of interest upon which to perform further sub-classification analyses.

Gray line: Non-classified states. This classification represents those points of the play session that do not fall under the three classifications made above. These ambiguous states of play may fall into one of the above categories when the tuning parameters for the classification algorithms are changed. They do not have to be directly considered in the analysis, but their presence can help researchers understand the granularity of their classification parameters.

The segments of time during which participants were engaged in coupled play (orange classification line) are a rich source of data. Those time segments can serve as a lens to go back to the initial sense-making curves to analyze the original sense-making curve values to see what types of activities participants were engaged in during the coupled interaction. For example, where both players simultaneously clamped on play actions? If so, were they mutually clamped on the same activity, such as attacking a castle? Another explanation could be one person was continuously adding meaning while the other person was continuously in the attentive waiting stage, which would result in the cumulative integral sum remaining largely the same and yielding a quasi-stationary state in the creative trajectory curve. The values from the original sense-making curve will reveal these nuanced details, allowing further sub-classification of the quasi-stationary play state into: dominant clamp (one person is adding new meaning and directing the play), mutual clamp (both participant are playing simultaneously), and rhythmic balanced interchange (participants are taking turns).

From this point in the analysis, it is also possible to count how much time was spent in each type of interaction trend and when each interaction trend tended to occur within the overall timeline of each pretend play session. With this data, it may be possible to identify successful collaboration strategies by comparing these broad interaction trends.

APPROACH LIMITATIONS

Without full validation, the findings from the CSM curve analysis are still in their preliminary phases. There are a few concrete activities that will help validate this technique that we are planning for future work. The primary concern with evaluating the CSM methodology is determining whether the classifications the technique generates (i.e. clamped/unclamped, mutual clamp, dominant clamp, etc.) correspond to those events in reality, i.e. does the theory correspond to the ground truth. One method for establishing ground truth would be to refer to the retrospective protocol analysis conducted with the participants to determine whether the remarks the participant made during a given time correspond to the cognitive state predicted by the theory. However, since the CSM analysis method was developed after the Drawing Apprentice and pretend play studies were designed and executed, the retrospective protocols did not focus on soliciting information explicitly related to the cognitive states later identified by the CSM framework.

There are also procedural limitations arising from the process by which analysts apply codes in the CSM approach. There is an inherent limitation in terms of the delay between an action onset and when a coder determines to move the coding slider to that corresponding value (since this is based on the analyst's reaction time and subjective judgment). Depending on each analysts' style, this delay may time-shift or change the granularity of certain events in the data, while still registering the overall deviation that is important later in the analysis. This limitation could drastically reduce inter-rater reliability while having an insignificant impact on the comparison of the actual sense-making curves and creative trajectories.

When users switch between codes using this tool, it is possible (and likely) that the system will sample data points during the process of moving from one code to another. This results in the system recording values between the current code and the target code. The number of values recorded between code switches increases depending on how far away the target code is from the current code since the user employs the up/down arrow keys to move between codes. This limitation of the tool creates some potential noise in the data as values are being recorded that were not intentionally coded by the analysts. However, this source of noise does not appear to significantly affect the overall reliability of the technique. One method to avoid this limitation is to use the number keys to control mode changes rather than the arrow keys.

Sense-making curves provide continuous data about interaction dynamics, but they do not contain contextual information about the type of actions that participants are engaging in through time or the conditions of their environment. To account for this important contextual information, the sense-making curve tool could allow users to associate text input (e.g. labels, tags) to different time segments of the sense-making curve. To reduce the amount of time it takes to tag and label events, the system could employ an autocomplete technique for typing tags as well as present the user with frequently used tags when they begin the event labeling task. This feature could combine the power of event-based coding (e.g. contextual details) with the continuous interaction analysis of the sense-making curve. With this information, the system could classify interaction dynamic trends as well as quantify the number of events and their temporal relationship.

DISCUSSION & FUTURE DIRECTIONS

Each sense-making curve represents a rich source of data that can be further analyzed using a variety of parameters and metadata that describe the features of each curve. Some initial metadata parameters to explore are the highest and lowest points of the curve, the number of inflection points, the magnitude of inflections, and the slope between inflections. Once each of the curves has been described using the metadata representation, the curves can be transformed to a new mathematical space for clustering analysis to reveal patterns and trends within and between different curves.

Graphically representing the creative trajectory curves and analyzing their trends can serve as a powerful data visualization and conceptual aid to reason about open-ended creative collaboration. The continuous nature of the sense-making curve data also provides opportunities to explore much more granular analysis and classification techniques. Further classification using the computational modeling convention mentioned in this paper (i.e. the stock market analysis technique) enables the possibilities of further quantification to facilitate comparisons throughout the dataset.

It might be possible to utilize machine learning algorithms to automatically code video data based on training sets comprised of human coded sense-making curve data paired with its video counterpart. The machine learning algorithm could attempt to code the video and inform the user about the segments of the curve in which it has less confidence about its coded value. The user could perform a targeted coding session to modify and update the low-confidence segments, thereby reducing the overall code application time while maintaining accuracy.

CONCLUSIONS

This paper presented the technical need and theoretical foundations for the proposed sense-making curve coding and analysis technique. The creative sense-making framework provides a new method to visualize and quantify the interaction dynamics of creative collaboration, e.g. the rhythm of interaction, style of turn taking, and manner in which participants are mutually making sense of a situation. The framework was applied to empirical studies of human collaboration (in the domain of pretend play) and AI-based systems (in the domain of collaborative drawing) to establish its validity through cross-domain application and inter-rater reliability within each domain. The theoretical justifications for the choice of the conventions for behavioral coding markers as well as graphical representation were described. We showed how the continuous nature of the sense-making curve data allows continuous functions, such as integrations, to provide quantitative information about how the collaboration is unfolding through time. This technique was applied to an exemplar data set to demonstrate the type of data that such an analysis yields. This technique may be generalizable to other fields studying and designing products for open-ended creative collaboration and co-creation.

ACKNOWLEDGEMENTS

Thanks to the analysts that helped apply and refine the creative sense-making approach: Margaret Hu, Yueting Lee, Shan Li, Lily Lau, and Margeaux Comerford. This work was supported by NSF Grant IIS-1320520.

REFERENCES

1. Malika Auvray, Charles Lenay, and John Stewart. 2009. Perceptual interactions in a minimalist virtual environment. *New ideas in psychology* 27, 1: 32–47.
2. Stefan Benus. 2009. Variability and stability in collaborative dialogues: turn-taking and filled pauses. In *Tenth Annual Conference of the International Speech Communication Association*.
3. Jerome Bourbousson and Marina Fortes-Bourbousson. 2016. How do Co-agents Actively Regulate their Collective Behavior States? *Frontiers in Psychology* 7.
4. Erin A Carroll and Celine Latulipe. 2012. Triangulating the personal creative experience: self-report, external judgments, and physiology. In *Proceedings of Graphics Interface 2012*, 53–60.
5. Erin A Carroll, Celine Latulipe, Richard Fung, and Michael Terry. 2009. Creativity factor evaluation: towards a standardized survey metric for creativity support. In *Proceedings of the seventh ACM conference on Creativity and cognition*, 127–136.
6. Sharolyn Converse. 1993. Shared mental models in expert team decision making. *Individual and group decision making: Current*, 1993: 221.
7. Nicholas Davis, Margeaux Comerford, Mikhail Jacob, Chih-Pin Hsiao, and Brian Magerko. 2015. An Enactive Characterization of Pretend Play. In

Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition, 275–284.
8. Nicholas Davis, Chih-Pin Hsiao, Kunwar Yashraj Singh, Lisa Li, and Brian Magerko. 2016. Empirically Studying Participatory Sense-Making in Abstract Drawing with a Co-Creative Cognitive Agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 196–207.
9. Hanne De Jaegher and Ezequiel Di Paolo. 2007. Participatory sense-making. *Phenomenology and the cognitive sciences* 6, 4: 485–507.
10. Hanne De Jaegher, Ezequiel Di Paolo, and Shaun Gallagher. 2010. Can social interaction constitute social cognition? *Trends in cognitive sciences* 14, 10: 441–447.
11. Ezequiel A Di Paolo, Marieke Rohde, and Hiroyuki Iizuka. 2008. Sensitivity to social contingency or stability of interaction? Modelling the dynamics of perceptual crossing. *New ideas in psychology* 26, 2: 278–294.
12. Valentina Fantasia, Hanne De Jaegher, and Alessandra Fasulo. 2014. We can work it out: an enactive look at cooperation. *Frontiers in psychology* 5.
13. Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological bulletin* 76, 5: 378.
14. Karl Friston. 2009. The free-energy principle: a rough guide to the brain? *Trends in cognitive sciences* 13, 7: 293–301.
15. Karl Friston. 2010. The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience* 11, 2: 127–138.
16. Karl Friston, James Kilner, and Lee Harrison. 2006. A free energy principle for the brain. *Journal of Physiology-Paris* 100, 1: 70–87.
17. Tom Froese, Hiroyuki Iizuka, and Takashi Ikegami. 2014. Embodied social interaction constitutes social cognition in pairs of humans: a minimalist virtual reality experiment. *Scientific reports* 4.
18. John S Gero. 2011. Fixation and commitment while designing and its measurement. *The Journal of Creative Behavior* 45, 2: 108–115.
19. Arthur M Glenberg. 1997. What memory is for: Creating meaning in the service of action. *Behavioral and brain sciences* 20, 1: 41–50.
20. Nitesh Goyal, Gilly Leshed, Dan Cosley, and Susan R Fussell. 2014. Effects of implicit sharing in collaborative analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 129–138.
21. Catholijn M Jonker, M Birna Van Riemsdijk, and Bas Vermeulen. 2011. Shared mental models. In *Coordination, organizations, institutions, and norms in agent systems vi*. Springer, 132–151.
22. Brigitte Jordan and Austin Henderson. 1995. Interaction analysis: Foundations and practice. *The journal of the learning sciences* 4, 1: 39–103.
23. Andruid Kerne, Steven M Smith, Eunyee Koh, Hyun Choi, and Ross Graeber. 2008. An experimental method for measuring the emergence of new ideas in information discovery. *Intl. Journal of Human–Computer Interaction* 24, 5: 460–477.
24. M Kim and Mary Lou Maher. 2005. Comparison of designers using a tangible user interface & graphical user interface and impact on spatial cognition. *Proc. Human Behaviour in Design* 5.
25. Harold A Linstone, Murray Turoff, and others. 1975. *The Delphi method: Techniques and applications*. Addison-Wesley Reading, MA.
26. Mary Lou Maher, Zafer Bilda, and Leman Figen Gül. 2006. Impact of collaborative virtual environments on design behaviour. In *Design computing and cognition'06*. Springer, 305–321.
27. Mary Lou Maher and Douglas H Fisher. 2012. Using AI to evaluate creative designs. In *DS 73-1 Proceedings of the 2nd International Conference on Design Creativity Volume 1*.
28. Etienne B Roesch, Slawomir J Nasuto, and J Mark Bishop. 2012. Emotion and anticipation in an enactive framework for cognition (response to Andy Clark). *Frontiers in psychology* 3.
29. R Keith Sawyer and Stacy DeZutter. 2009. Distributed creativity: How collective creations emerge from collaboration. *Psychology of Aesthetics, Creativity, and the Arts* 3, 2: 81.
30. Philipp Schwartenbeck, Thomas FitzGerald, Ray Dolan, and Karl Friston. 2013. Exploration, novelty, surprise, and free energy minimization. *Frontiers in psychology* 4: 710.
31. Sune Steffensen and Sarah Bro Pedersen. 2014. Temporal dynamics in human interaction. *Cybernetics & Human Knowing* 21, 1–2: 80–97.
32. Sune Vork Steffensen. 2012. Care and conversing in dialogical systems. *Language Sciences* 34, 5: 513–531.
33. John Robert Stewart, Olivier Gapenne, and Ezequiel A Di Paolo. 2010. *Enaction: Toward a new paradigm for cognitive science*. MIT Press.
34. David Vernon. 2010. Enaction as a conceptual framework for developmental cognitive robotics. *Paladyn, Journal of Behavioral Robotics* 1, 2: 89–98.
35. David Vernon. 2014. *Artificial cognitive systems: A primer*. MIT Press.
36. Pawel Wozniak, Nitesh Goyal, Przemyslaw Kucharski, Lars Lischke, Sven Mayer, and Morten Fjeld. 2016. RAMPARTS: Supporting Sensemaking with Spatially-Aware Mobile Interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2447–2460.