Creativity Metrics for a Lead-and-Follow Dynamic in an Improvisational Dance Agent

Meha Kumar, Duri Long, Brian Magerko
The Expressive Machinery Lab
Georgia Institute of Technology
Atlanta, GA 30332 USA
{mkumar77, duri, magerko}@gatech.edu

Abstract

Being creative in movement-based improvisational environments, such as dance floors, poses a difficult challenge for computers. It is particularly challenging for computers to judge the creativity of inputs and responses in these open-ended domains. *LuminAI* is an improvisational dance installation where an artificially intelligent agent dances with a user. In this paper, we enable the agent to judge the creativity of dance gestures through the development of creativity metrics, including the *novelty*, *value*, and *surprise* of a gesture. We then use these metrics to implement a lead-and-follow dynamic: when the user is performing less creatively, the agent tries to lead by performing more creatively. A user study is performed to compare the original and lead-and-follow systems, with results showing users found the lead-and-follow system's gestures lower in quality than the original, more surprising than the original, and found the original system more engaging and influential on their actions.

Introduction

Though natural for most humans, being creative in movement-based improvisational environments, such as sports games and dance floors, poses a difficult challenge for computers. Agents with the ability to improvise movement could be used to inspire human choreographers, create more engaging training sequences in sports or physical therapy, or generate less predictable non-player character actions in video games.

One way to develop creative artificially intelligent (AI) agents is to equip them with an understanding of what is considered "creative". Algorithms can measure the creativity of an artifact, such as a painting, poem, or dance move. The AI then uses this metric to guide itself as it explores the conceptual space, resulting in the generation of more creative artifacts.

Creativity can be defined in many ways, but one of the leading definitions used in the computational creativity community describes a composite score of three parts: the *novelty*, *value*, and *unexpectedness* of an artifact (Maher 2010; Boden 2004). Previous research has provided a basis for evaluating artifacts of creative systems (Maher 2010; Maher and Fisher 2012). However, in real-time interactive environments, efficiency is of utmost importance, so the metrics need to be adapted to produce creative results while not disrupting the experience.

One example of a research project where creativity metrics have been implemented in a real-time interaction is the Robot Improv Circus. The Robot Improv Circus is an installation in the domain of improvisational acting, where an agent plays an improv theater game with a user (Jacob 2019). Improvisational acting, however, requires general knowledge about the world to create an engaging performance. The difficult task of amassing knowledge for agents to use in human-agent interactions, or the *knowledge-authoring bottleneck*, becomes more evident as the number of possible meaningful actions increases. Since the space of improv theater gestures is essentially any action a human could perform in real life, the agent struggled to produce meaningful acts.

Comparatively, the domain of dance is simpler to interpret with computation, because it is more abstract (Jacob et al. 2013). Therefore, creativity metrics for an improvisational AI dance agent can be developed as a stepping stone towards a creative dance agent.

In this work, we will discuss the development of creativity metrics for an interactive dance installation called *LuminAI* (Figure 1). We present our implementation of these metrics, which uses the three-pronged definition of creativity described earlier to judge gestures.

Evaluating the creativity metrics can indicate whether they are useful in *LuminAI* and potentially other projects. Previous research with the Robot Improv Circus has shown that direct attempts to evaluate creativity metrics can be tricky (Jacob 2019). In this experiment, for each metric, participants were shown pairs of improv gestures and asked to select which one they thought would score higher for that metric. The results showed the developed metrics did not match human ratings. However, it also showed that the participants' selections were close to random– meaning that experiments asking humans to directly rate gestures based on novelty, value, and surprise may be inherently flawed. Jacob (2019) suggests that having people compare two gestures without the greater context of an interaction may be too challenging. In an effort to determine whether the metrics are useful without directly asking about them, we have used the metrics to develop a potentially more engaging version of *LuminAI*.

Figure 1: The *LuminAI* installation. The black figure on the screen is the agent, while the pink figure is the user's "shadow."

Specifically, we have created a lead-and-follow dynamic between the user and agent. When the human is performing less creatively, the agent tries to lead by performing more creatively, and vice versa. Finally, we will present the results from a preliminary user study investigating whether the metrics are noticeable when the lead-and-follow dynamic is switched on, whether the agent seems more creative when leading, and whether the lead-and-follow dynamic is more engaging for users. By focusing on the effects on the users' experiences with the new system, we will gain insights into the effectiveness of the creativity metrics.

The *LuminAI* System

LuminAI is an improvisational art installation in which the user can dance with an artificially intelligent agent, shown as a humanoid figure on a screen. The system uses a Microsoft Kinect™ to sense a user's movements, using periods of stillness to segment motion into discrete gestures. In this study, we utilize a discrete gesture mode, where the agent does not dance while the user is performing, and no user input is recorded while the agent is performing. The agent learns from its users by storing user gestures into a database. The agent currently responds with the following response modes (Jacob et al. 2013; Jacob and Magerko 2015):

- Mimic: agent repeats the user's gesture
- Transform: agent alters the user's gesture (e.g. swap the movements of the arms)
- Random Recall: agent pulls a random gesture from its database
- Related-gesture Recall: agent pulls a gesture from its database that is similar in calculated key aspects (such as *energy*, *size*, and *tempo*)

Related Work

Defining creativity

Boden (2004) defines creativity in three parts: *novelty* (how new the artifact is to the agent), *unexpectedness* (how surprising the artifact is in the current context), and *value* (the quality of the artifact). This definition has been widely explored within the computational creativity community. All three components are necessary when trying to design an AI agent which produces creative artifacts without human input. Agents which use only novelty require human input to guide them away from producing noise–in effect, the humans provide the metric of value (Kar, Konar, and Chakraborty 2015). Unexpectedness serves to account for artifacts which may have been seen before, but surprise us when presented in the current context.

Researchers have taken these three criteria and formalized them mathematically (Wiggins 2006), proposed possible algorithms to calculate these formalizations (Maher 2010; Lehman and Stanley 2011; Maher and Fisher 2012), and implemented such algorithms in attempts to make creative agents (Jacob 2019). Another proposed element of creativity is *typicality* (how well the artifact conforms to the expectations of its domain) (Ritchie 2007); however, Jacob (2019) argues this metric is accounted for so long as the system does not maximize unexpectedness and novelty above all else. The metric of value is present to prevent this, and thus typicality would not provide any additional information. We use Boden's definition in this work and develop a three-component algorithm for creativity based on novelty, value, and unexpectedness (referred to from now on as *surprise*).

Detailing each metric

Novelty Novelty can be defined as how different an artifact is from other artifacts within the same domain that the observer has seen in the past (Boden 2004). Computationally, artifacts can be represented as vectors within some space, where the dimensionality of the space is determined by the number of features of the artifacts. Novelty is then intuitively understood as how far away an artifact is from all others, using any appropriate measure of distance.

One approach to determine novelty is to cluster the artifacts and determine to what degree the new artifact matches the nearest cluster (Maher and Fisher 2012; Barto, Mirolli, and Baldassarre 2013). Another approach utilizes Self-Organizing Maps in conjunction with clustering, which also reduces the dimensionality of the data and can provide useful visualizations (Maher 2010; Maher and Fisher 2012). A third approach involves determining the average distance from the artifact to its K-Nearest Neighbors, where K is some empirically defined natural number (Lehman and Stanley 2011; Maher and Fisher 2012).

In *LuminAI*, the gestures are represented as motion-capture data tracking each joint over time. This is an extremely high-dimensional representation, so any approach will require dimensionality reduction first. An effective pipeline for dimensionality reduction in *LuminAI* has already been produced (Liu et al. 2019). This pipeline relies on *feature reduction* and Principal Components Analysis to compress the data to a few (typically 2 or 3) dimensions. Feature reduction refers to the condensing of a gesture's original motion-capture representation into a smaller representation which does not include every single recorded frame. Specifically, the pipeline calculates the 15 most representative frames in any particular gesture, referred to as the *keyframes* of a gesture, and stores only those frames.

The reduction produced by this pipeline loosely groups gestures based on the major body parts involved in the movements (e.g. leg and hip movements are grouped together, left arm movements are grouped together). After viewing this reduced data in a visualization tool, it became clear that clustering is not an effective basis for measuring novelty because the data is sparse and the clusters are not separated enough for the valid novelty calculations. Therefore, we use average distance to the K-Nearest Neighbors on the reduced data as a measure of novelty.

Implementation The gestures, represented as high-dimensional motion capture data, are reduced to two dimensions using a modification of the previously built dimensionality reduction pipeline described in Related Work (Liu et al. 2019). The modification focuses on the selection of the 15 keyframes. Rather than calculate the most representative keyframes of a gesture, which was inefficient, keyframes are chosen at uniform intervals from the gesture. This modification changed the chosen keyframes only slightly, leading to approximately the same reduction. The reduction seemed to preserve the pipeline's ability to group gestures loosely based on which body parts were involved, while running much faster.

In this space, we calculate the novelty of a gesture as the average distance to its K-Nearest Neighbors (with K=5 yielding the best spread of novelties). The K-Nearest Neighbors algorithm utilizes an R-Tree of known gestures to quickly find neighbors. However, this novelty score would not be entirely useful for programming: average distances can only be compared to one another, and thresholds cannot be set to distinguish "high" and "low" scores. In order to scale the average distances to $[0, 1]$, we pass them through an *adaptive scaling* tool. This tool dynamically sets the highest and lowest values it has seen thus far, allowing future numbers to be scaled to $[0, 1]$ based on these values. This gives a final novelty score.

To avoid getting extreme novelty scores (i.e., 0 or 1) as the scaling tool sees its first few distances, a preprocessing step is needed. On startup, the system calculates the novelty of all gestures in the database, thereby passing all the unscaled scores through the scaling tool. If the database is reasonably varied, future gestures should not exceed the bounds set by the scaling tool too often.

Value

Background Value can be defined as the usefulness, performance, and quality of an artifact to the observer, in the context of the surrounding culture. Clearly, value is highly dependent on the domain of artifacts being considered. We look at a definition of value that can be applied to any domain before focusing on value in dance gestures for *LuminAI*.

Maher and Fisher (2012) describe measuring value as similar to measuring novelty, that is, as a distance: this time in a "performance space." In their application of measuring the creativity of laptop designs, they identified features relevant to the value of a laptop design by hand, creating vectors within a performance space. They then used a distance-to-centroid measure to determine overall value: laptops farther from the centroid of all the laptops were considered higher in value. Ideally, the agent will be able to determine the value of an artifact without human assistance. Therefore, we need to allow the agent to distill a dance gesture vector into its relevant features automatically.

The essential question here is: what features makes a dance gesture valuable? Montero (2012) suggests that the observer's experience performing a particular dance style makes them a better judge of the quality of a movement. However, the agent in *LuminAI* has no perception of how it "feels" to perform a movement, so we measure value solely based on the agent's visual perceptions.

Researchers have tried to understand why certain dancers are better than others through the lens of attractiveness (Neave et al. 2010; McCarty et al. 2017); they suggest that the motivation behind perceptions of "good" and "bad" dancing is reproductive. For example, McCarty et al. (2017) identify greater hip movement as one quality of a good female dancer which may indicate female mobility. One possible method of determining the quality of a gesture, therefore, is quantifying the amount of movement in certain key body parts. An aggregate score for quality may then be obtained from the various key body parts.

To avoid basing the value metric on attractiveness, which may introduce notions of gender and sexuality to the project, we also turn to a popular framework for understanding motion called Laban Movement Analysis (LMA) (Laban and Ullmann 1971). This framework was developed primarily for performers themselves, but also lends itself well to computational analysis. LMA interprets movement with four aspects:

- Body: what each part of the body is doing and how body parts are related and connected
- Effort: the qualities of the movement such as flow and weight
- Shape: the overall shape of the body and how it changes
- Space: the movement's interaction with the surrounding environment

By using select LMA aspects as the relevant features for value, the agent could interpret movement using the same metrics that many human dancers use. A complication is that each style of dance may require its own calibration of the aspects. Since those interacting with the system in this study are novices who are not specialized in any style of dance, we will not use LMA to tune *LuminAI*'s value metric to any one style of dance. Instead, we will use the attraction-based definition of quality dance gestures, mitigating potential biases as much as possible. Future work could explore the use of LMA for specific dance styles.

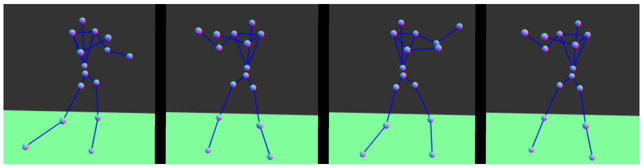

Figure 2: A gesture with high value.

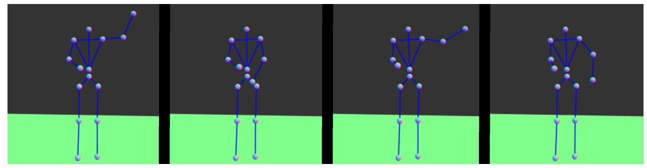

Figure 3: A gesture with low value.

Implementation Following the definition of a quality dance move as based on attractiveness (Neave et al. 2010; McCarty et al. 2017), certain key indicators of good dance moves from both men and women are measured. By measuring the indicators for both men and women outlined by Neave et al. (2010) and McCarty et al. (2017), we hope to mitigate bias towards any particular gender when evaluating value. Specifically, we measure the amount of hip movement, shoulder movement, asymmetrical thigh movement, and asymmetrical arm movement from the motion capture data.

In order to measure these quantities, we determine the average amount of movement between consecutive frames. For efficiency, we first reduce the frame rate by half by removing every other frame. For hip movement, we calculate how far the left and right hip joints have moved between each pair of consecutive frames. These distances are summed to get a total amount of hip movement in the gesture. Then, we divide this by the number of frames in the gesture to achieve the average hip movement per frame. The average ensures that long gestures are not higher in value than shorter ones simply because they accumulate more movement. For shoulder movement, the same method is used, but we track the left shoulder, right shoulder, and neck joints instead.

For the asymmetrical thigh movement, we find the vectors representing the change in position for the right and left knee joints between each pair of consecutive frames. The left and right change vectors are subtracted to find a vector representing the asymmetrical movement between frames. We take the magnitude of this vector as our measure. As with hip movement, we sum this measure across all pairs of consecutive frames and divide by the total number of frames to achieve an average amount of asymmetrical thigh movement in the gesture. For asymmetrical arm movement, the same method is used, but we track the left and right elbows and wrists instead.

Once all four measures are found, we sum them together to get a total amount of valuable movement. As with novelty, this score needs to be scaled to $[0, 1]$ to be useful. An adaptive scaling tool and preprocessing step are used in the same manner. Gestures marked with high and low values are shown in Figure 2 and Figure 3, respectively.

Surprise

Background Surprise can be defined as the unexpectedness of an artifact based on the observer's expectations. As opposed to novelty, an artifact can still be considered surprising even if it has been seen before; surprise takes into account recent events which shape an expectation.

Barto, Mirolli, and Baldassarre (2013) define two ways to quantify surprise: 1) surprise as deviation from a prediction, and 2) surprise as the degree of difference between the agent's beliefs before and after an event. The latter may be quantified using Kullback-Leibler divergence (Barto, Mirolli, and Baldassarre 2013). The difference between a predicted gesture and the observed gesture may be computed using a distance in the feature space, similar to novelty. Maher and Fisher (2012) consider an artifact surprising when it creates a new cluster within the conceptual space; this artifact changes the agent's model of expectation significantly. In another work, Maher (2010) notes that surprise occurs when the observer has established a pattern, which the artifact then violates.

We utilize the definition of surprise as deviation from an expected dance gesture. We build expectations based on the last movement performed by the human and/or agent. In other words, we aim to find a dance move surprising if it deviates from the prediction built by the previous dance move.

Implementation Following the definition of surprise as deviation from an expectation (Barto, Mirolli, and Baldassarre 2013), we must first define what the expected response gesture is when some dance move is performed. In order to truly know the expected response, a large data set would need to be collected consisting of many gestures with their observed response gestures. This would be a direction for future work, but for the initial development of the surprise metric in this paper, we make the assumption that the least surprising response would be mimicry. The next expected gesture, then, is the same as the current gesture.

Now, the deviation from the expected dance gesture can be defined. As Jacob (2013) describes, each gesture in LuminAI has certain key values associated with it, based on theories of movement, such as the *energy*, *size*, and *tempo* of the movement. We define surprise using the difference between the expected and actual gestures in two aspects: one to account for the difference in key values, and one to account for the difference in which body parts are used in the gestures. We use both because one can imagine a gesture that is surprising in only one aspect or the other. If a gesture consisting of large, fast arm circles followed a gesture of small, slow arm circles, it would be surprising (although the same body part was used). If a gesture of big, sudden kicks followed a gesture of big, sudden punches, it would be surprising (although the gestures have similar key values).

To obtain the part based on the key values, the difference between the expected and actual gestures' key values are summed. This value is passed through an adaptive scaling tool. To obtain the part based on body parts, we reduce the dimensions of both the expected and actual gestures using the same dimensionality reduction pipeline used in the novelty calculation. In this reduced space, the distance between the two gestures is measured. This value is passed through another adaptive scaling tool. We chose to use separate scaling tools for the key value and body part components because the values they produce may be on wildly different scales, and we do not want one measure to dwarf the other when summed. The key values component and the body parts component are then summed and passed through a third adaptive scaling tool to obtain a final score for surprise.

Lead-and-Follow Interaction

Background

Leading and following may create a more interesting user experience as the agent no longer solely responds to the user; it can also directly attempt to inspire the user. Lead-and-follow dance agents have been developed in the past with major limitations. Berman and James (2015) proposed a dance agent which dances with higher or lower intensity in response to its human partner. Our work is distinct in that the agent is less limited in its possible gestures, the agent uses creativity metrics as the basis for judging movement, and most importantly, the agent is able to both lead and follow the exchange. This may help the agent and human be true equals in the interaction, leading to a more stimulating experience for the user.

In the past, a lead-and-follow agent was built in *LuminAI* (Winston and Magerko 2017). This version of the *LuminAI* agent judged when to lead or follow based on the enthusiasm of a gesture (how wide or high-tempo it is) rather than the creativity of a gesture. The study found that users could tell the difference between the lead-and-follow and original versions, but that the original version was preferred. By using creativity metrics, we plan to build on this work by developing a lead-and-follow dynamic that is more engaging for users than the one implemented by Winston.

Implementation

Leading is added as an additional response mode (adding on to the existing modes of mimicking, random gesture recall, transformation, and related-gesture recall (Jacob and Magerko 2015)). The LuminAI agent selects from these response modes every time a user gesture is detected; it does not take into account whether it had been leading previously. Future work could explore incentivizing staying in leading mode for several turns.

First, the *LuminAI* dance agent must decide whether to lead based on the user's gesture. If the measured novelty, value, or surprise of the gesture is low, it tries to lead. During development, the thresholds for leading were tuned until the agent led when the user performed known gestures, and followed when the user performed new or interesting gestures. Concretely, the agent leads when the novelty, value, or surprise scores of the user's gesture are below 0.5. When leading, the agent selects a response gesture which has high surprise in the context of the user's gesture. The agent finds the ten gestures with highest surprise from the database and randomly selects one to perform.

Ideally, the agent would perform the gesture with the highest value from the set of ten, but this response usually returned the same gesture repeatedly. This could be because gestures which measure high in value may be further away from other gestures in both the spaces used to calculate surprise (described earlier). Then, the same high-value gesture will be one of the ten highest-surprise gestures, and it will be returned every time the agent tries to lead. When not leading, the agent falls back on its existing response modes.

Evaluation

Methodology

In order to evaluate the system, a preliminary user study was performed. Subjects interacted with both the original *LuminAI* system and the creative lead-and-follow system described in this paper (referred to here as *C-LuminAI*) and provided feedback in the form of a survey and interview. Both systems were set to the discrete gesture mode described earlier, where only one party is dancing at a time. This ensures that the last gesture performed by either party is always known, which eliminates confusion about which gesture the dancers are responding to. Both systems' agents were allowed to perform one gesture per turn. Both systems were pre-loaded with the same database of gestures. There were about 30 gestures, all recorded by researchers who are novice dancers. The systems were presented to subjects in a randomized order. Both interactions were video recorded.

First, subjects were allowed to familiarize themselves with the agent for a few turns, until the subject could successfully complete a gesture and see the agent's response. Then, subjects interacted with each agent for about five minutes. Following the interactions, the users filled out a survey which asked about their perceptions of the agent's dance moves, the levels of creative contribution and control both parties had over the interaction, their preferred agent, and the subject's experience and comfort with dance. 5-point Likert scales were used to ask subjects about qualities of each system (e.g. "The dance moves the agent performed were good." was asked using two Likert scales, one for each agent).

A short interview was then conducted to collect qualitative descriptions of the differences between the agents and reasoning behind the participant's preferred agent.

Results

Seven subjects participated in the study, all of whom were college students with varying experience and comfort dancing. The small sample size means the results are not statistically significant, but they show some interesting preliminary trends. The results are shown in Figure 4. Overall, the survey data showed that *LuminAI* performed better gestures than *C-LuminAI*, *LuminAI*'s gestures were less surprising than *C-LuminAI*'s, *LuminAI* engaged participants better than *C-LuminAI*, and *C-LuminAI* was not influenced by users as easily nor influenced users as much as *LuminAI*. The survey data also showed that users felt they contributed more to

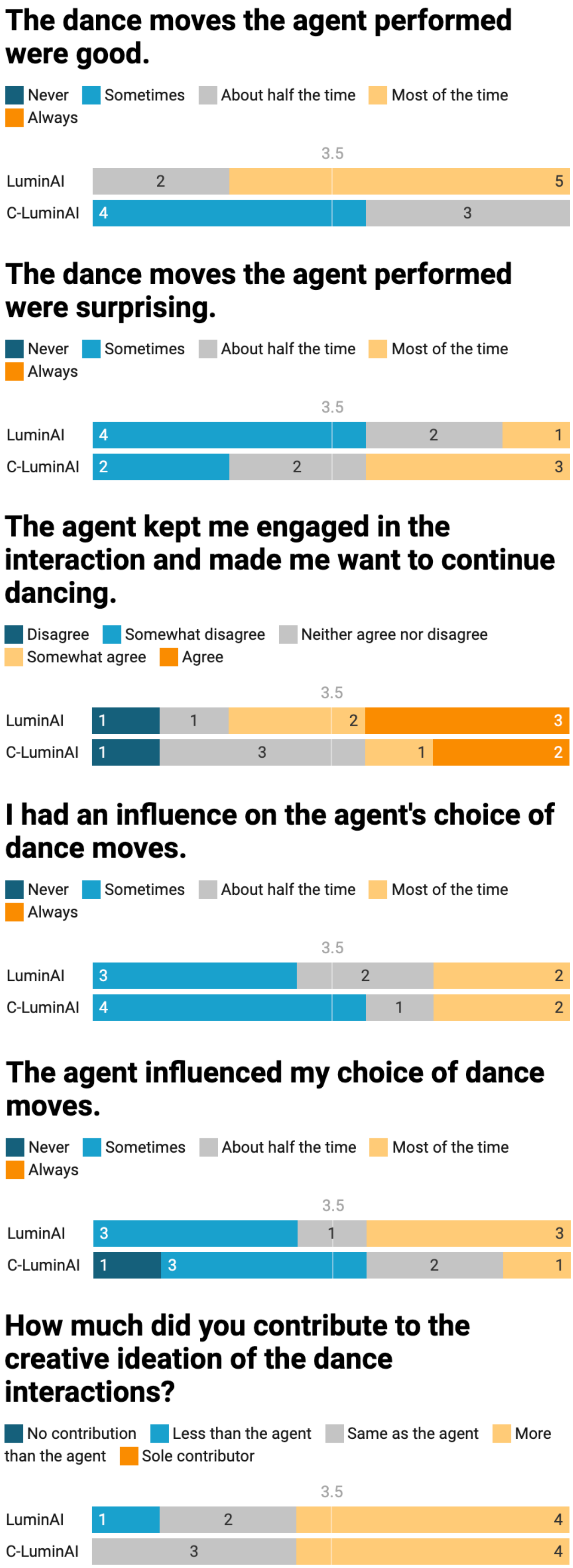

Figure 4: Survey results. The halfway point of 3.5 responses is marked. The number in each color bar shows how many participants selected that response.

the creative ideation of the interaction than the agent in both systems. Neither system was discernibly preferred overall. One subject stated in the interview that ”neither one seemed like they were responding, just doing what they wanted.”

The interview and video data showed that *C-LuminAI* performed shorter, more inhuman movements which were more varied than *LuminAI*'s. Six out of the seven participants contrasted *C-LuminAI*'s "short", "jerky", or "inhuman" movements with *LuminAI*'s "longer", more "human", more "natural", or "better" movements. Two participants said *LuminAI* made them want to mimic the agent's gestures because the gestures were more realistic. One participant said that *LuminAI* could better "mimic and complement" their gestures, and another said *LuminAI* seemed to "pick up [their] new moves". Three participants mentioned that *LuminAI* was "predictable" or "repetitive", though one of these three said the repetition made it more human-like. One participant said *C-LuminAI*'s "weird moves made [them] want to be wilder".

Discussion

The strongest result pointed to *C-LuminAI* giving nonhuman movements. This is likely because while leading, the agent was choosing among the most surprising gestures it could find, without taking the value of those gestures into account. This may have also led it to choose short gestures more often than *LuminAI*: the response modes of mimicking and transforming the user's move are chosen more often in *LuminAI* and are guaranteed to be similar in length to the user's gesture. That *C-LuminAI* gave strange gestures is actually promising when paired with the result that users found *C-LuminAI* more surprising. However, the agent seems to have veered too far off the surprising end of gestures into ones which limited the user's ability to respond. Users felt that *LuminAI* was able to influence their behavior and engage them more with its better quality gestures. Interestingly, this did not correlate to *LuminAI* being preferred overall.

In Winston and Magerko's (2017) study of their lead-and-follow agent in *LuminAI*, users preferred the original agent. However, the users' reasoning for preferring one system to the other in that study were based on the increased mimicry of the original agent, while in our study, users' reasoning was more based on the quality and variety of gestures performed. Both studies found some comments suggesting users enjoy the agent mimicking them. Based on these findings, a clearer lead-and-follow dynamic could allow users to gain the satisfaction of having the agent mimic them while the agent follows. Then, perhaps when the agent is leading, users may be more receptive to and less disappointed by the agent's moves.

If the lead-and-follow dynamic is made clearer, and the gestures chosen when the agent is leading made more valuable, then the agent may reap the benefits of both versions. The lead-and-follow dynamic could be made clearer using text prompts, visual highlighting of which party is leading, and only using mimic and transform response modes when following.

The gestures chosen when leading were strange because *C-LuminAI*'s agent chooses one gesture randomly from the top ten most surprising gestures, but it ought to factor in value. This could be enabled by 1) vastly expanding the database of gestures so there are many high-value gestures that will often also be high-surprise when compared to the user's gesture, or 2) changing the dimensionality reduction technique used on gestures so high-value gestures are not far away from all other gestures. This may not be possible, because better gestures may be inherently different from other ones. Either of these solutions would allow the leading mode to choose a high-value gesture from the top ten most surprising gestures, instead of randomly selecting one of the ten. The agent would then be able to be varied like *C-LuminAI* when leading, but always have realistic gestures.

Overall, one of the most important factors in the user's perception of this system is the quality of the gestures performed. The value metric described in this paper is a powerful tool for controlling the quality of gestures that are played back (or perhaps even stored into memory). One important addition would be factoring in the length of a gesture to the value metric. In addition, the surprise metric seems to successfully deliver more surprising gestures based on the survey results.

Future Work

Going forward, the developments described in this paper can be used to make *LuminAI* a more engaging system and to help explore the creative potential of the agent. The lead-and-follow dynamic developed in this paper can be improved and incorporated into *LuminAI* to reduce the repetition of gestures and provide an interesting new response mode. The creativity metrics can be used to explore *creative arcs*, as described by Jacob (2019). Creative arcs are paths the agent takes over its performance based on the creativity metrics: for example, the agent may start performing with high value and progress to low value, while also progressing from low novelty to high novelty. Overall, the ability of the agent to autonomously judge gestures for their creativity opens the door for new reasoning strategies and gesture selection algorithms.

Conclusion

In this work, algorithms were developed to measure the creativity of a dance gesture in the improvisational dance installation *LuminAI*. These metrics judged the novelty, surprise, and value of a gesture. These metrics can be used in the future to control the quality of gestures performed by the agent and add new reasoning strategies. The lead-and-follow dynamic developed in this paper improved the variety of gestures performed by the agent, but decreased their quality. Further development of this dynamic could increase user engagement with the system or help explore creative relationships between machine and human collaborators.

Acknowledgements

The authors would like to thank Lucas Liu, Cassandra Naomi, Swar Gujrania, and Jonathan Moon for their helpful discussions and continued work on the LuminAI system.

References

Barto, A.; Mirolli, M.; and Baldassarre, G. 2013. Novelty or surprise? *Frontiers in psychology* 4:907.

Berman, A., and James, V. 2015. Kinetic dialogues: Enhancing creativity in dance. In *Proceedings of the 2nd International Workshop on Movement and Computing*, 80–83. ACM.

Boden, M. 2004. In a nutshell. *The Creative Mind* 1–24.

Jacob, M., and Magerko, B. 2015. Interaction-based authoring for scalable co-creative agents. In *ICCC*, 236–243.

Jacob, M.; Coisne, G.; Gupta, A.; Sysoev, I.; Verma, G. G.; and Magerko, B. 2013. Viewpoints AI. In *Ninth Artificial Intelligence and Interactive Digital Entertainment Conference*.

Jacob, M. 2019. *Improvisational Artificial Intelligence for Embodied Co-creativity,*. Ph.D. Dissertation, Georgia Institute of Technology.

Kar, R.; Konar, A.; and Chakraborty, A. 2015. Dance composition using microsoft kinect. In *Transactions on Computational Science XXV*. Springer. 20–34.

Laban, R., and Ullmann, L. 1971. The mastery of movement.

Lehman, J., and Stanley, K. O. 2011. Novelty search and the problem with objectives. In *Genetic programming theory and practice IX*. Springer. 37–56.

Liu, L.; Long, D.; Gujrania, S.; and Magerko, B. 2019. Learning movement through human-computer co-creative improvisation. In *Proceedings of the 6th International Conference on Movement and Computing*, 5. ACM.

Maher, M. L., and Fisher, D. H. 2012. Using AI to evaluate creative designs. In *DS 73-1 Proceedings of the 2nd International Conference on Design Creativity Volume 1*.

Maher, M. L. 2010. Evaluating creativity in humans, computers, and collectively intelligent systems. In *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*, 22–28. Desire Network.

McCarty, K.; Darwin, H.; Cornelissen, P. L.; Saxton, T. K.; Tovée, M. J.; Caplan, N.; and Neave, N. 2017. Optimal asymmetry and other motion parameters that characterise high-quality female dance. *Scientific reports* 7:42435.

Montero, B. 2012. Practice makes perfect: the effect of dance training on the aesthetic judge. *Phenomenology and the Cognitive Sciences* 11(1):59–68.

Neave, N.; McCarty, K.; Freynik, J.; Caplan, N.; Hönekopp, J.; and Fink, B. 2010. Male dance moves that catch a woman's eye. *Biology letters* 7(2):221–224.

Ritchie, G. 2007. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines* 17(1):67–99.

Wiggins, G. A. 2006. Searching for computational creativity. *New Generation Computing* 24(3):209–222.

Winston, L., and Magerko, B. 2017. Turn-taking with improvisational co-creative agents. In *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*.