COMPUTATIONAL CREATIVITY EVALUATION

Anna Kantosalo
OUTLINE

WHY TO EVALUATE
WHEN TO EVALUATE
WHAT TO EVALUATE
WHO SHOULD EVALUATE
HOW TO EVALUATE
WHY TO EVALUATE

"A comparative, scientific evaluation of creativity is essential for progress in computational creativity, not least to justify how creative a computational creativity system actually is."

- Jordanous, 2012
WHY TO EVALUATE

"A comparative, scientific evaluation of creativity is essential for progress in computational creativity, not least to justify how creative a computational creativity system actually is."

- Jordanous, 2012

- Evaluation highlights progress
- Evaluation shows what can be improved
- Evaluation (when done well) allows for comparison with other systems
- Evaluation argues how a system is creative
WHEN TO EVALUATE

- Evaluation should ideally be a part of every project undertaken in CC
- Nowadays some type of evaluation is also mandatory for publication!
- In addition to evaluation done within the system, the system should be evaluated in a comprehensive way at multiple stages during its development
- Evaluation is an iterative, on-going process
WHEN TO EVALUATE

- Systems should be evaluated when
  - A project starts: What can be achieved with the chosen methodology; setting future evaluation targets?
  - Once a part of the project is finished: Does the system do what it is intended to do; how can we boost its performance?
  - Once the whole project is finished: Does the system as a whole do what it was intended to do; how can we boost its performance? And how does the system compare to other similar systems?
WHEN TO EVALUATE

- Systems should be evaluated when
  - A project starts: What can be achieved with the chosen methodology; setting future evaluation targets?
  - Once a part of the project is finished: Does the system do what it is intended to do; how can we boost its performance?
  - Once the whole project is finished: Does the system as a whole do what it was intended to do; how can we boost its performance? And how does the system compare to other similar systems?

Summative evaluation provides a summary of a system’s creativity while formative evaluation provides constructive feedback on its strengths and weaknesses.
WHAT TO EVALUATE
JORDANOUS’ FOUR PPPERSPECTIVES ON COMPUTATIONAL CREATIVITY

- Person/Producer
  - Qualities of the system producing creative artefacts
  - (Could also apply to whoever designs and implements the system)
- Process
  - Algorithmic processes within, and interactions with the creative entity
- Product
  - The result of the creative process
- Press/Environment
  - The environment in which the creativity is situated
WHAT TO EVALUATE

EVALUATION CRITERIA

  – Suggested as metrics for evaluating the Product
    – Jordanous suggests can be used to evaluate all Ps
  – Suggests computing ratings for different criteria, e.g. the average typicality of produced items
Colton (2008): The creative tripod

- Skillfull, Appreciative, Imaginative
  - Colton originally suggested a shift from evaluating the product to evaluating the producer
  - He also recognizes that the programmer, the computer and the consumer can all contribute skill, appreciation and imagination to the creative experience
  - Definition of criteria is vague, but one interpretation is
    - Skillfull – Ability to produce
    - Appreciative – Ability to evaluate the value of the product
    - Imaginative – Ability to produce novel items
WHAT TO EVALUATE
EVALUATION CRITERIA

- Colton, Charnley and Pease (2011): Computational Creativity Theory
- IDEA: Well-being rating, Cognitive-effort rating
  - Pease and Colton then again shifted the focus from viewing the producer, process or product to viewing the effect the creative act has on an ideal audience
  - Well-being rating: the personal hedonistic value of a creative act
  - Cognitive-effort rating: the time a person is prepared to spend interpreting the creative act and its results
  - In the IDEA model, these two ratings are used to compute various effects for a creative act, e.g. disgust
WHAT TO EVALUATE

EVALUATION CRITERIA

- Jordanous (2012): Components of creativity
- 14 themes identified from literature

- Active involvement and persistence;
- Dealing with uncertainty;
- Domain competence,
- General intellect;
- Generation of results,
- Independence and freedom,
- Intention and emotional involvement,

- Originality,
- Progression and development,
- Social interaction and communication;
- Spontaneity/Subconscious processing;
- Thinking and evaluation;
- Value;
- Variety, Divergence and experimentation

- Can be used as evaluation criteria selectively
WHAT TO EVALUATE

EVALUATION CRITERIA

- Van der Velde et al. (2015): Originality, Emotional value, Novelty/innovation, Intelligence, Skill
- A fresh look to evaluating products
- Intended for outside evaluators
WHO SHOULD EVALUATE

- **Creator vs. Audience**
  - Should the system be evaluated by the system’s creators themselves, by outside experts, or by the intended audience

- **Experts vs. Laymen**
  - Should the system be evaluated by experts of computational creativity, field specific experts, peers, or laymen

- Evaluation can move on multiple levels
- Different targets can be evaluated by different persons
- A combination can be used to achieve more holistic and more useful results
HOW TO EVALUATE
STANDARDISED PROCEDURE FOR EVALUATING CREATIVE SYSTEMS

SPECS: A Standardised Procedure for Evaluating Creative Systems was proposed by Jordanous (2012) as a domain independent way to define an evaluation process for a creative system.
HOW TO EVALUATE
STANDARDISED PROCEDURE FOR EVALUATING CREATIVE SYSTEMS

- **Step 1: Defining creativity**
  - A definition the system should satisfy to be considered creative
  - What does it mean to be creative in general?
  - What aspects of creativity are important in the particular domain of the system?
  - What are you going to evaluate? Which Ps are interesting to you?
HOW TO EVALUATE
STANDARDISED PROCEDURE FOR EVALUATING CREATIVE SYSTEMS

- Step 2: Identifying Strands (which of the Ps) to Test for
  - Transform your definitions from step 1 to standards for testing the system
  - E.g. see evaluation criteria reported before
HOW TO EVALUATE
STANDARDISED PROCEDURE FOR EVALUATING CREATIVE SYSTEMS

- **Step 3: Testing Systems**
  - Test creative system against the standards set in step 2 and report the results
  - Tests depend on standards, preferences, capabilities, equipment and facilities of the researchers involved
  - Methods can be quantitative and qualitative
  - Suitable evaluators should be selected
CONCLUSIONS

- Evaluation is critical to examine the creativity of computational creativity systems
- Evaluation is an essential requirement of good research
- To conduct a good and thorough evaluation, the researcher must identify when to evaluate, what to evaluate, who should evaluate and how to conduct the evaluation
REFERENCES

• Anna Jordanous (2016) Four PPPPerspectives on computational creativity in theory and in practice, Connection Science 28:2, 194-216


• Simon Colton, John Charnley, and Alison Pease (2011). Computational creativity theory: The FACE and IDEA Descriptive models, In the Proceedings of the 2nd International Conference on Computational Creativity 2011, pp.90-95

• Simon Colton (2008). Creativity versus the perception of creativity in computational systems. Proceedings of AAAI symposium on creative systems (pp. 14-20), Stanford, California, USA