The Profile of A Creative Program

Graeme Ritchie
University of Aberdeen

Profound questions (about "strong" creativity):

> Can a computer program be truly creative?

> What are the most creative acts performed by any computer program?

> What computational mechanisms are most effective in producing creativity?

(and others)

How can these become empirically testable?

These depend on the more precise question:

➢ How has this computer program performed when it produces results?

Even if we are just interested in "weak" creativity, we still want to know this.

This talk:

Suggests some aspects of a "creative$_w$" program's behaviour which might be relevant to assessing its success (and its "creativity$_s$").

That is, what might we look at in order to get a more precise idea of how well a "creative" program is doing?"

- Just the abstract ideas – no details of testing methods, statistics, etc.

- NOT "measuring" creativity , but "profiling" the program's behaviour.

Some central ideas

Success: Is the program managing to create acceptable artefacts?

[So we need to be able to tell if output is/is not an "acceptable" artefact.]

Novelty: How similar are the created artefacts to previously known exemplars?

[So we need some notion of "known", and of "similar".]

Quality: How good are the created artefacts?

[So we need a measure of the "value" of artefacts.]

Making it more precise

The program has produced some output, a set of *result items*.

There is an intended class of artefacts, represented as a fuzzy set defined by a rating on some scale (e.g. [0,1]); the *Typicality* measure. (cf. Wiggins' R)

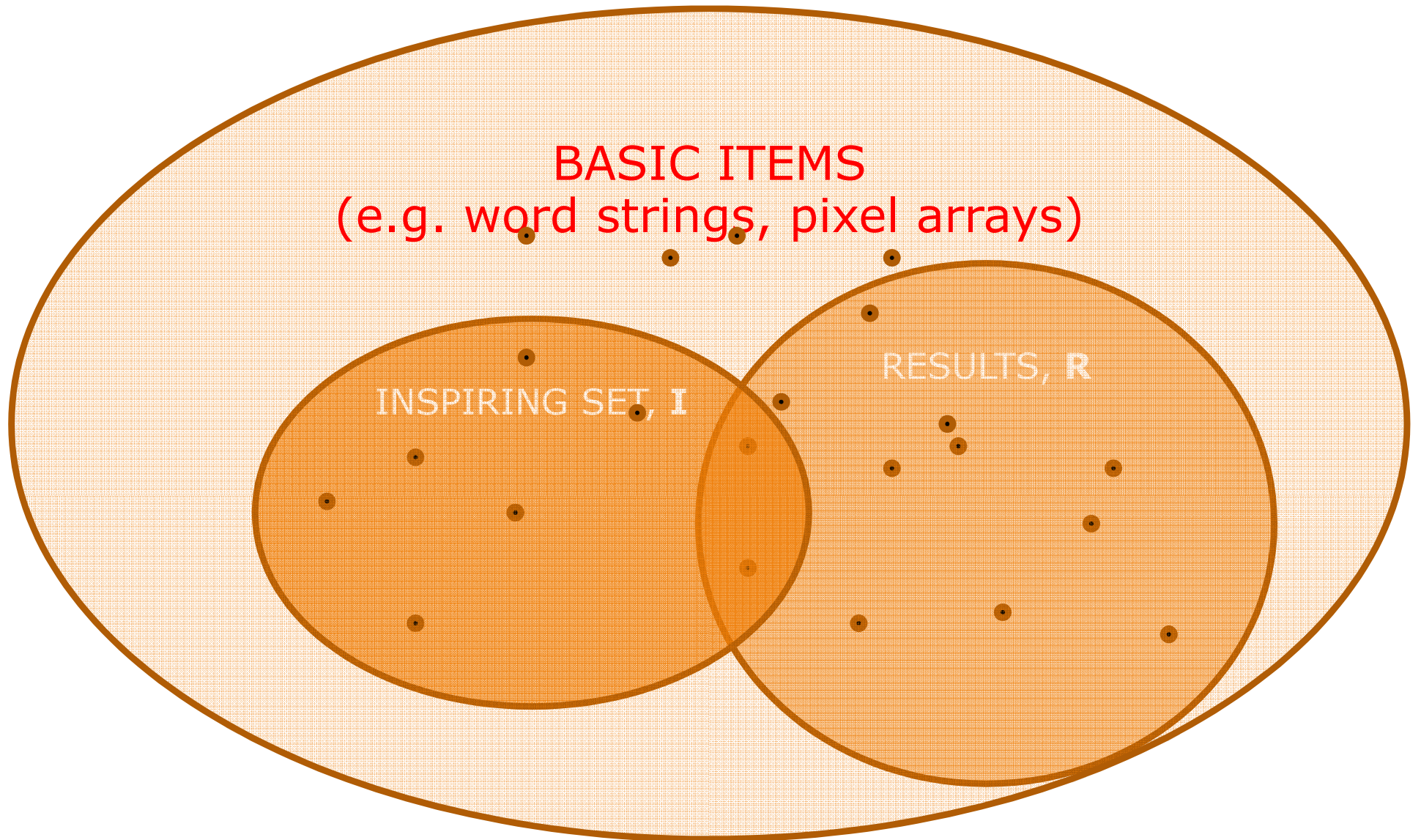The program design is based on some known *exemplars*: *The Inspiring Set*.

There is some measure of *similarity*, scoring the distance/difference between items.

There is a measure of how "good" an artefact is, defined by a rating on some scale (e.g. [0,1]); the *Value* measure. (cf. Wiggins' E)

BASIC ITEMS
(e.g. word strings, pixel arrays)

INSPIRING SET, **I**

RESULTS, **R**

Making it more precise

The program has produced some output, a set of *result items*.

There is an intended class of artefacts, represented as a fuzzy set defined by a rating on some scale (e.g. [0,1]); the *Typicality* measure.

The program design is based on some known *exemplars*: *The Inspiring Set*.

There is some measure of *similarity*, scoring the distance/difference between items.

There is a measure of how "good" an artefact is, defined by a rating on some scale (e.g. [0,1]); the *Value* measure.

Can the program replicate known exemplars? ["*reverse engineering*"]

Maybe not very "creative", but tests the accuracy of the abstract model.

Of particular interest: those exemplars on which the program's design was based : the "inspiring set".

What proportion of the "inspiring set" appear in the results?

To what extent are the program's outputs really examples of the intended class of artefacts? (e.g. its poems are indeed poems, its melodies are really melodies.)

What is the *average* rate of "typicality" in the result set?

What *proportion* of the result set are at a high level of "typicality" ?

How novel is the program's output?

A major factor in "strong" creativity: originality, innovation, difference from previous artefacts.

What *proportion* of the inspiring set are very similar to items in the results?

What *proportion* of the results are very similar to items in the inspiring set?

Novelty and Typicality

Novelty alone is uninteresting, unless the artefact is a real exemplar.

Do those results which are not similar to the inspiring set have a high level of typicality, on *average*?

What *proportion* of the results are both not similar to the inspiring set and at a high level of typicality?

What *proportion* of the results which are not similar to the inspiring set are at a high level of typicality?

How good are the program's outputs?

One of the factors in creativity ("weak" or "strong") – the "quality" of the artefacts produced.

> **What is the *average* "value" of items in the result set?**

> **What *proportion* of the result set are at a high level of "value" ?**

Novelty and Quality

Combining innovation with high quality is very interesting.

Do those results which are not similar to the inspiring set have a high level of value, on *average*?

What *proportion* of the results are both not similar to the inspiring set and at a high level of value?

What *proportion* of the results which are not similar to the inspiring set are at a high level of value?

Relationship between typicality & quality - (1)

A competent human creator produces artefacts which:
- lie firmly within the intended genre (i.e. achieve high typicality)
- are of good quality (i.e. achieve high value).

What *proportion* of the highly typical items in the result set are at a high level of "value" ?

Relationship between typicality & quality – (2).

A *very* creative human produces artefacts which:
- are unusual (i.e. not very typical)
- are of high value (good quality)

> **What *proportion* of the results are both UNtypical items and at a high level of "value" ?**

> **What *proportion* of the results which are UNtypical items are at a high level of "value" ?**

> **What *proportion* of the high value results are UNtypical items?**

Novelty and Typicality and Quality

The right combination of these attributes is central in creativity.

> **What *proportion* of the results which not similar to the inspiring set are both typical items and at a high level of "value" ?**

> **What *proportion* of the results which not similar to the inspiring set are both UNtypical items and at a high level of "value" ?**

So far:

We can characterise several facets of a creative system's behaviour using a relatively small number of constructs.

These can be made precise.

ratio($V_{\gamma,1}$ ($R - N_\delta(I)$) $\cap$ $T_{0,\beta}$ ($R - N_\delta(I)$), ($R - N_\delta(I)$)

where:

R is the results set

I is the inspiring set

$N_\delta(I)$ is the set of items within a distance of items in I

$V_{\beta,1}(X)$ is the subset of X whose value is between β and 1

$T_{0,\alpha}(X)$ is the subset of X whose typicality is between 0 and α.

So far:

We can characterise several facets of a creative system's behaviour using a relatively small number of constructs.

These can be made precise.

But some quantities might be hard to determine (e.g. the inspiring set).

Any more?

"Fine tuning" (Colton, Pease, Ritchie 2001)

To what extent has the knowledge used in a program been specially crafted (by the designer) to produce the results which occurred?

(i.e. to replicate the "inspiring set")

Suppose the following is the case:

- we can segment the knowledge base K of the system into parts, $K_1,....K_n$;
- we can run the program with one specific K_i removed;
- we can compute the difference between the results found with and without the use of K_i (i.e. those results which rely on K_i);
- we can compare these selected results with the inspiring set.

Compute the ratio of these selected results which are in the inspiring set to those which are not in the inspiring set.

The bigger this is, the more K_i has contributed to "fine-tuning".

Aggregate the scores for all the different K_i (e.g. by averaging); this gives a "fine-tuning" rating for K.

These are *not* a checklist to see if a program is being truly ("strongly") creative.
   -- they are "probes" to clarify how a program is/is not behaving

Not all factors have been considered here; e.g.
  -- human intervention
  -- surprise (is it different from untypicality?)

These are not a definitive closed set of criteria.
  -- please devise some more.

They illustrate that some of these ideas can be made precise.
  -- we need precise definitions of program behaviour if we want to test them empirically.

The story so far….

- Ritchie (2001) suggests definitions for 14 of these criteria.

- Gervas(2002) uses these to assess the WASP poetry generator, and raises some criticisms.

- Pereira et al (2005) apply the criteria to two programs: Divago concept generator, Dupond paraphrase generator.

- Haenen & Rauchas (2006) apply the criteria to a melody generator.

- Ritchie (2007) reviews all the above and expands list to 18 criteria.

Concerns

- These "criteria" are NOT a checklist to measure "strong" creativity.

- Keeping the formulae but changing the meaning of the components obscures the issue.

Where now?

- Non-arbitrary ways of defining these constructs in measurable ways.

- Methodologies for actual testing.

- Similar "probes" for the internal workings of the program?

The End

(or perhaps merely the beginning….)

Assessing Creativity
Graeme Ritchie (2001).
Proceedings of AISB Symposium on AI and Creativity in Art and Science, Society for the Study of Artificial Intelligence and Simulation of Behaviour.

Some empirical criteria for attributing creativity to a computer program
Graeme Ritchie (2007)
Minds and Machines 17 (1), pp. 67-99.

The Effect of Input Knowledge on Creativity
Simon Colton, Alison Pease, Graeme Ritchie (2001)
In *Proceedings of Workshop on Creative Systems: Approaches to Creativity in AI and Cognitive Science*, Fourth International Conference on Case-Based Reasoning. Vancouver.