What if computers had ideas of their own?

Knowledge, Computational Creativity and the What-If Machine

Mark Granroth-Wilding

May 21st 2015
KVIT, Linköping

This document presents a talk given at KVIT 2015, to an audience largely made up of cognitive scientists. It's on the subject of Computational Creativity (CC), a subfield of AI probably not familiar to anyone who was there. The talk is therefore in two halves. In the first half, I introduce CC, a young and rapidly growing field, and try to give some idea of why it's an exciting topic to be researching and why now is an exciting time to be doing it. In the second half, I illustrate my points with specific reference to the project I'm currently working on – the *What-If Machine*, or *WHIM*.

Contents

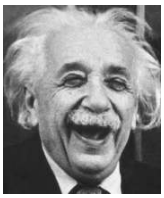
1 Computational Creativity

1.1 Introduction

Why creativity?

- Disinctively human activity
- Central to culture, science, philosophy, ...
- Important to understanding human intelligence
- Not well understood

> *Creativity is intelligence having fun*
>
> — *Albert Einstein*



Why should we be interested in studying creativity at all? Here are a few reasons.

It's considered to be a distinctively human activity, or even, depending on what type of creativity you're talking about, perhaps even uniquely human. It's well accepted as being something that lies at the heart of human culture, science, philosophy, and so on.

It's clear, then, that it's a very important part of the working of human intelligence (individual and communal), so understanding it really ought to be something that cognitive scientists are interested in.

What is more, there are a lot of interesting unanswered questions regarding creativity. What are the cognitive processes that underlie it? What sort of knowledge or reasoning capacities does it rely on? What makes some people more successful than others at being creative in particular ways? And so on...

Why **computational** creativity?

- Same as most AI:
  - Way to understand human behaviour better
  - Precise, testable theories
  - Useful[1] tools, automating human tasks
- Specific to CC:
  - Contribution to philosophical questions
  - System take on parts of creative process
  - Human-machine artistic collaboration

[1] and/or exciting and/or lucrative

So, why should anyone be interested in CC, then?

Some of the reasons are rather similar to reasons for a cognitive scientist to study any branch of AI. It gives us a means of trying to understand the human mind and human culture better through a process of building **concrete models** that we can **implement**, **experiment** with and, if we're lucky, draw some conclusion about the theory from.

There's also, of course, the possibility that the models work well, in which case they may lead to **useful tools** for automating human tasks, interacting better with humans, etc.

But there are also some motivations more specific to CC. The computational approach can make a contribution to **unanswered philosophical questions** about nature of creativity, and this is something that's already happened to some degree.

Creative systems can take on **parts of creative processes** that involve humans as well. This might be doing the **mundane bits** that we don't want to do ourselves, or the **difficult bits**, leaving us to concentrate on the parts that interest us. Or they might provide us with **new sources of inspiration** for our own creativity. There's even the possibility of devising **human-machine collaborations** on artistic projects, producing creative processes that simply wouldn't have been possible otherwise.

All of these things call for systems that can interpret creative artefacts in a human-like way and, in some cases, produce creative responses to them.

> The philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative.
>
> — *Colton & Wiggins (2012)*, Computational Creativity: The Final Frontier?

This is a standard way to begin any talk on CC! It's a *working definition* of the field proposed by two of the major active figures in the community, Simon Colton and Geraint Wiggins. It's been much discussed, both leading up to its publication and since.

It's worth unpacking a little here, because it gives a neat indication of what the community's interested in.

Philosophy:
There are lots of open philosophical questions. This is an interesting aspect of the field, but I'll largely pass over it here. It's worth noting, though, that the formal and practical work in CC interacts with philosophy and leads to new perspectives on the questions.

Science:
CC encourages us to construct theories of creativity that can be precisely defined and empirically tested. This can make a **contribution to cognitive science** in much the same way that computational linguistics, for example, has done, and for the same reasons.

Engineering:
CC has lots of exciting applications in systems in which software takes on **some creative responsibility**. Note that we're not necessarily interested in complete, end-to-end systems where the machine does everything. Many of the examples I give of applications involve automating some part of a creative process. Or we might prefer to think of tools that assist humans in creative tasks.

An interesting thing to notice about this definition is that it avoids saying **what creativity is**! This is not an accident. Experience has shown that where we try to define is a precise manner what 'creativity' means in order to do it, we invariably end up with something that doesn't look anything like what humans actually recognise as creativity! We can make progress in CC without having a precise definition and here it's left up to the humans to decide what is and isn't creative.

Also worth noting is that it doesn't constrain us to **Turing test-style evaluation**, where the aim is to convince a human that the system is another human. There are many reasons why this is a bad idea, which I won't go into here. This definition is open to, for example, humans knowing that the artist is software and interacting with it on those terms.

1.2 Applications

Applications

- Short story generator
- Assistant for writers
- Interactive educational tools for children
- Live game generation (storylines)
- Targeted advertisting
- Automatic contextual/responsive music generation
- Collaborative partners for human artists

These are just a few examples of applications of CC. Some of them are extensively explored already, others still open research topics.

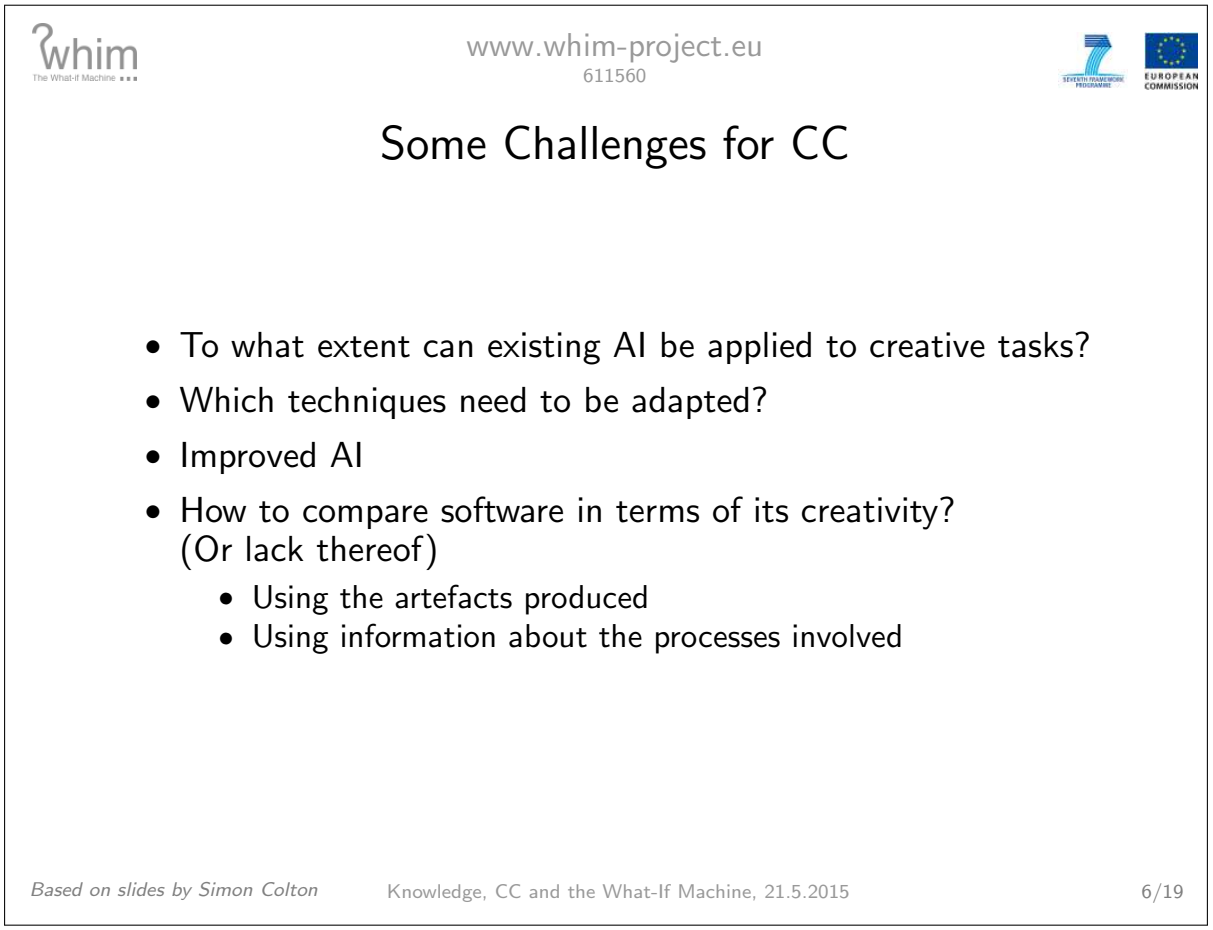
This is a small selection – no doubt you can think of more!

I do not plan to discuss...

- What *is* creativity?
- Is it inherently human?
- Could a computer ever be *truly* creative?
- Isn't it really the *programmer* who's being creative?
- What's the point in creativity if no human is involved?
- Are you trying to put artists out of work?

These are some questions I don't plan to address now, largely at the philosophical end of the field. Some of them are interesting, all are frequently discussed. But none are directly relevant to the work I'm involved in or other work I'll discuss. It's worth noting that the science and engineering side of CC can make progress without first having full answers to these sorts of questions.

1.3 Challenges and contributions



Here are a few of the general challenges currently being tackled by the field.

You might imagine that creative interpretation and generation can be in part addressed by **existing AI techniques**. To what extent is this true? Where do they fall down? In other words, what in particular makes these creative tasks have a **different set of requirements** for a computational system to traditional AI task?

Then, how can we **adapt or augment existing techniques** to the creative domain?

A less obvious question is: how could work in CC lead to **improvements in tasks already tackled by AI**? Are there tasks that we suddenly see new ways of tackling when we think of them as creative and bring to bear on them some of the advances in CC?

Evaluation is one of the most difficult problems for any CC project. Traditional measures of success often simply don't apply. For example, measuring the **accuracy** of a system's output compared to human answers makes no sense in a task where there is no right or wrong answer. Instead, we have to find other means to compare systems on the basis of their output.

But, note that that's not the only way to approach evaluation. We might be interested not just in a system's output, but also in the **process** that produced it. That maybe sounds like an odd idea, but humans assessing some artistic work are often more interested in the process or thoughts behind its creation than the artefact itself.

Some contributions: applied

- Mathematical discovery
  - automated theory formation in pure mathematics
  - tasks in number theory/algebras
  - combination of automated reasoning systems

- Visual arts
  - automated painter: graphics and AI
  - emotion detection drives painting
  - scene construction, collage generation

- Video game design
  - games which dynamically adapt to players
  - models of user enjoyment/immersion
  - generation of topical arcade games

The field is young, but there have already been some significant contributions. Here are a few brief examples.

Mathematical discovery is a highly creative process – for example, the process of coming up with new theorems in pure maths, the problem tackled by the HR system. Here, as in CC more generally, we need to provide systems for exploration and discovery in **spaces that are unknown**, or at least ill-defined, before we start.

Here's an example of automating artistic creativity – the Painting Fool, an **automated painter**. This is an example of somewhere we can use existing techniques from AI and computer graphics, but, of course, this open-ended task calls for more. The system makes use of **emotion detection** to motivate its decisions.

Video games are a great domain for CC, providing us with lots of opportunity for **dynamic, responsive generation** of whole worlds, sensitive to the player's state, the state of the game, or other factors, like topical issues. Angelina is a system that tries out these sorts of possibilities.

Some contributions: philosophical/formal

- Rejection of the Turing test
- Developed notions of intentionality, appreciation and imagination in software
- New descriptive theoretical models
  - Exploring models through implementation
- Challenged use of created artefacts alone to assess creativity

Philosophical discussions occupy a lot of research time in CC and there are many unanswered questions – it's a young field. I said I wouldn't talk much about philosophy, but here are a couple of examples. These sorts of contributions have served to guide work in the community.

One example is the round **rejection of Turing test**-style tasks for evaluation. There are many reasons for this, which I'll not go into here.

Many **theoretical notions** have been discussed and developed in the new context of **creative software** – ideas like imagination.

A key idea I've already talked about is taking a computational approach to building **theoretical models**: we can explore new models by implementing them and refining the theory as we draw conclusions from the implementation. There's a lot of work along these lines and I will later discuss one example in detail.

Aside: Typicality

- Usually try to model typicality & expectation
- Why is this useful for *creative* generation?

- Most of a creative artefact not novel
- Model norms to know how to diverge from them meaningfully

Before I go any further, let me make a brief aside. The usual set of techniques we use in AI/machine learning are about modelling **typical or common properties** of whatever it is we're modelling – language, images, semantics, etc. It might seem that this is inappropriate once we're building creative systems: surely these properties are irrelevant when we're generating new, creative output?

There are two key reasons why this is not true. Firstly, novelty in creative works is generally carefully packed in material conforming to expectations and it is often from this that gains its effectiveness. Consider, for example, this Magritte painting: most of it is entirely dull and conforming to expectations, and it is this that makes the unexpected aspect right in the middle so surprising!

Secondly, in order to be able to diverge from norms in a meaningful way, the system needs a model of what they are to start with. It can then knowingly violate them in particular (and carefully limited) ways. Picasso, for example, knew exactly what norms and conventions he was challenging in his art. In contrast, the painting to the right was produced by a toddler throwing paint.

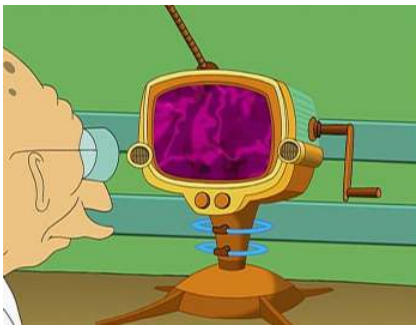
2 WHIM

2.1 Introduction

www.whim-project.eu
611560

The What-if Machine

- Simple creative fictional ideas
- May take *"what if...?"* form
- At the heart of many stories, paintings, etc



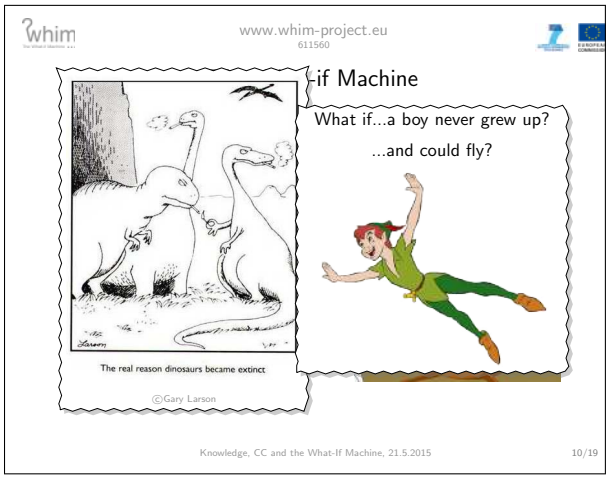
Knowledge, CC and the What-If Machine, 21.5.2015 10/19

I'm now going to give the specific example of the project that I'm currently involved in – the What-If Machine (WHIM). It's an EU project, a collaboration between five universities across Europe.

Many creative ideas take the form of a small fictional alteration to the world. These sorts of ideas can often be expressed as a "what-if...?" sentence. Production of this sort of hypothetical or counter-factual scenario is a common part of everyday life and many creative artefacts – books, films, games, paintings, adverts, etc. – can be thought of as deriving from such an idea.

Consider this Far Side cartoon: *the real reason dinosaurs became extinct*. Getting the joke involves working out the implicit *what if* underlying it.

Then, here's another what-if scenario, that might suggest a familiar story.



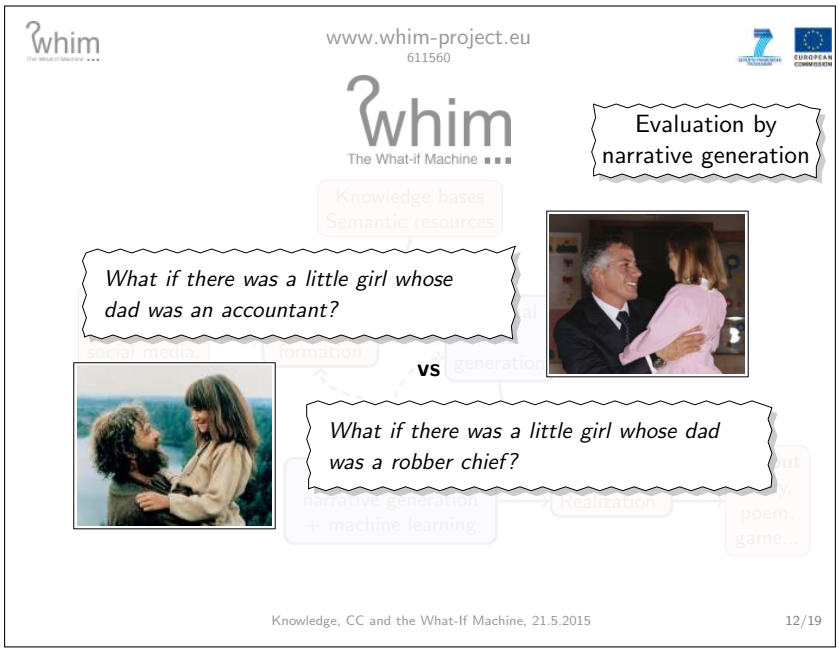
WHIM: some questions

- What makes ideas interesting?
- Can we generate/select them automatically?
- What resources are required? (knowledge, reasoning, ...)
- What processes must be involved in producing/evaluating them?
- Can we make use of automatically generated ideas?

This leads to some interesting questions about this type of idea.

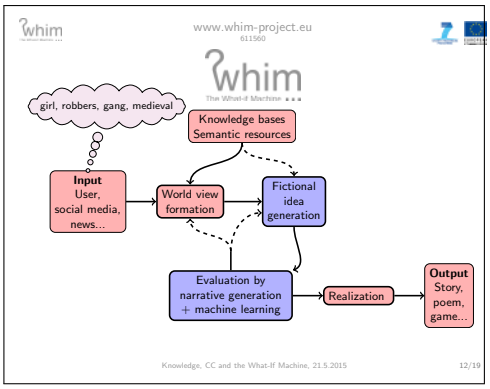
One way to get partial answers to some of these is to propose a plausible architecture for process of generating what-if ideas, try implementing it computationally and then see where it fails – it may well not be what we expect! We can then examine why it failed and get some insight into what's required for success. What is more, to the extent it does succeed, it has practical uses.

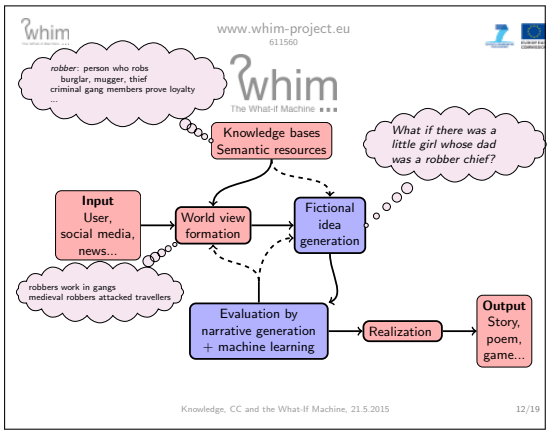
2.2 Architecture



So, we're investigating a particular computational model of one specific kind of creativity, namely that that lies behind this type of fictional idea. The key idea of the project is the evaluation of ideas by **narrative generation**. Consider these examples: why is the first idea uninteresting, whilst the other seems a lot more promising?

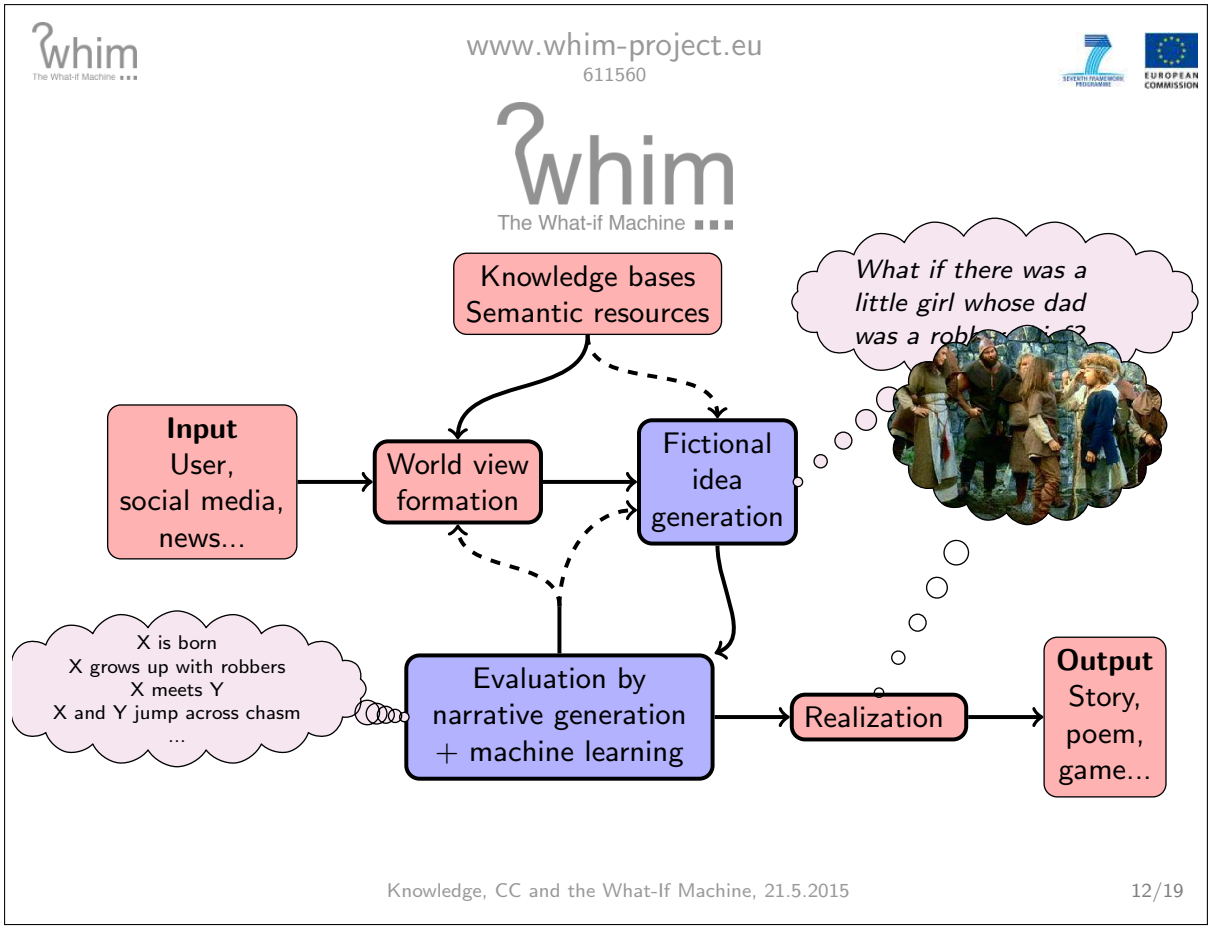
In the WHIM model, a system generates ideas and evaluates how good they are by trying to generate narratives from them. The system's architecture looks broadly like this (right). On the basis of some input, it builds a *world view* – some collection of world knowledge about concepts it considers relevant. This may draw on existing knowledge bases or information extraction techniques.





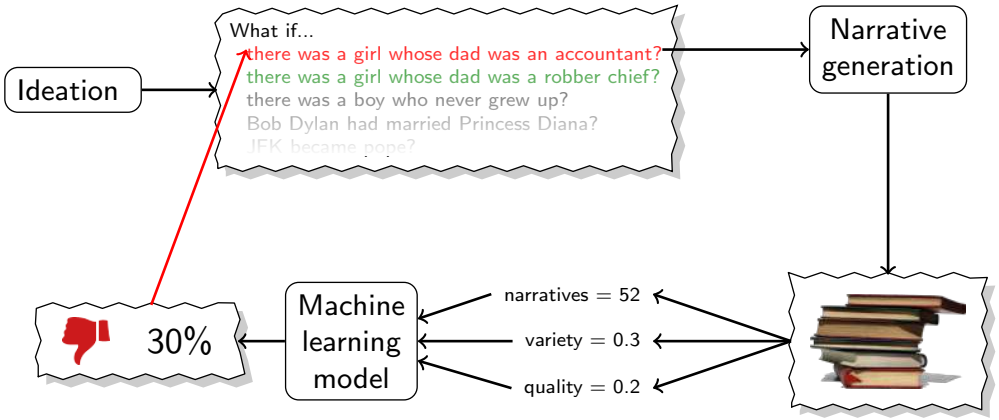
It then uses this to generate lots of fictional ideas – the process known as *ideation*. It tests out each idea by trying to generate narratives and running some features based on the results through a machine learning model, which decides how good each idea is. At this stage, it might go back to the world view building step, or just request more ideas, until it's happy with one.

Finally, the satisfactory idea is output in some form in which it can be evaluated by humans. This could be just a simple what-if sentence, or it could be something more elaborate, like a story, a game or a picture.



This process calls for a large amount of world knowledge of different sorts, and this is a key problem that we're tackling. What's clear already is that the generation of ideas itself is easy! For example, you can make minimally informed changes to known facts and produce lots of mediocre what-ifs, with some good ones. Evaluating ideas by narrative generation is hard, partly because narrative generation itself is a hard problem!

Don't get *any* ideas...

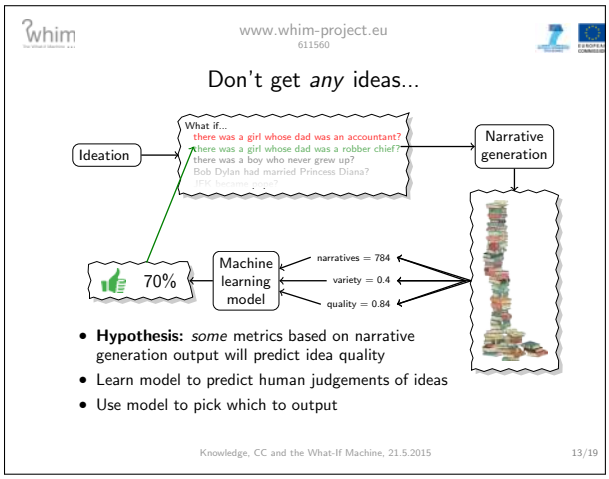


- **Hypothesis:** *some* metrics based on narrative generation output will predict idea quality
- Learn model to predict human judgements of ideas
- Use model to pick which to output

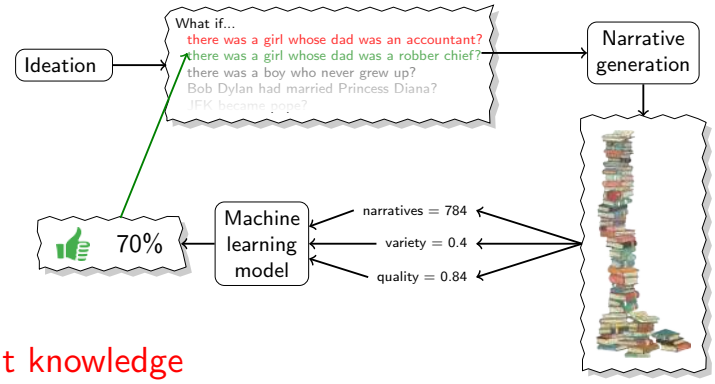
The main hypothesis, then, is that we can use *some* metrics based on narrative generation output to make good predictions of human judgements of idea quality.

The ideation engine produces lots of ideas with only simplistic constraints. We generate narratives from then and produce metrics over generated narratives. These could include, for example, the number of narratives, variety or some sort of quality measure. We're not committed to particular metrics: the idea is to simply output lots for now. We learn a machine-learning model to predict overall idea quality score from the metrics.

The output of this tells us whether the idea's worth pursuing. Note that this is a supervised learning task: we'll be collecting human ratings of ideas to train the model on.



Some requirements



- Event knowledge
- Relations, facts, world knowledge
- Narrative devices
- Plausibility measures

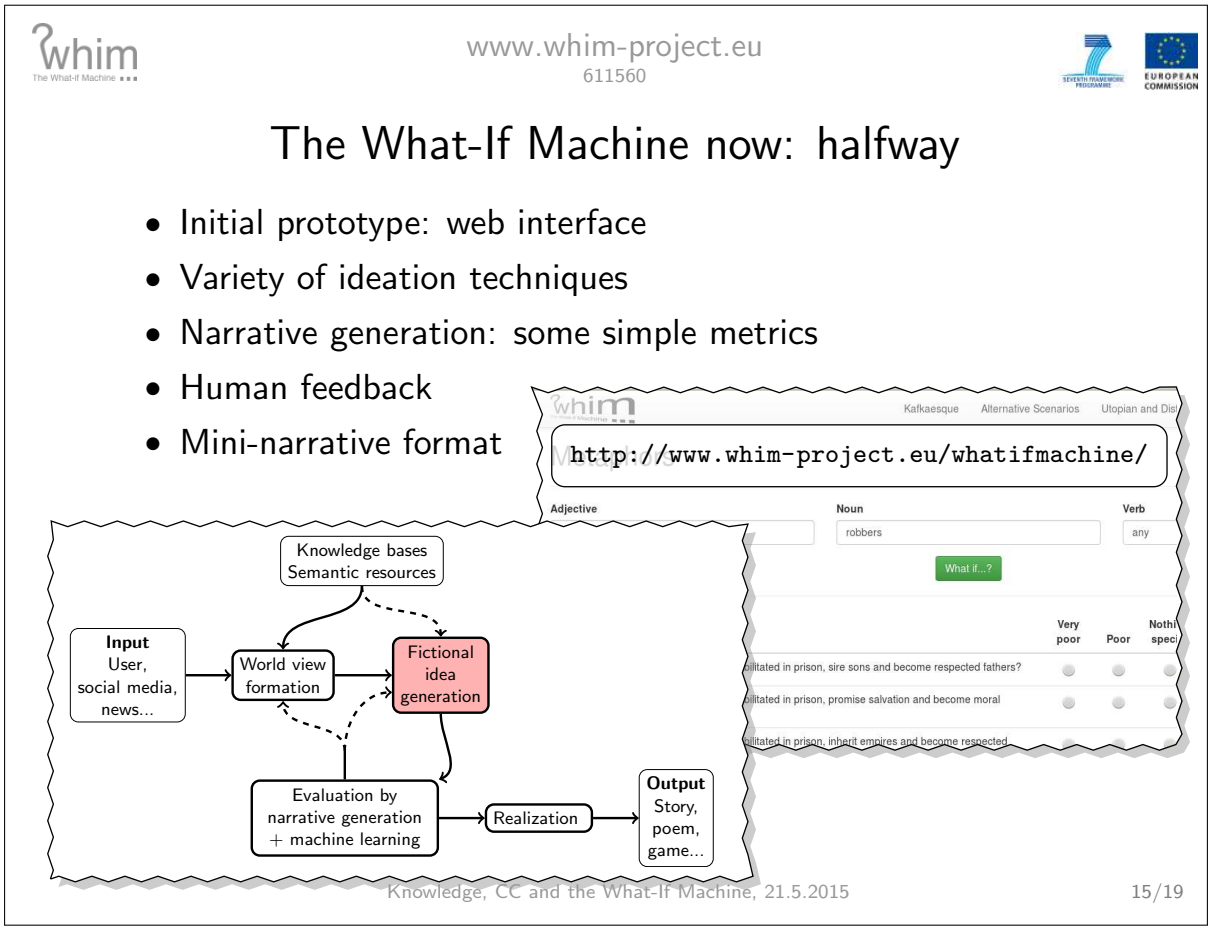
Crucially, we need to be able to expand an idea into narratives following from it to get these metrics. One of our partner teams on the project are specialists in narrative generation and will use their existing armoury of techniques.

Several types of knowledge are important for this process

- Events: expectations about sequences, consequences, causality
  Bill drops glass → *glass breaks*, but also,
  Bill orders food → *Bill drinks*
- Relations between entities, facts about things, world knowledge
  Bill Clinton was president of the US
  Cows eat grass
- Narrative devices, tropes
  Someone goes missing ⇒ *Villain deceives victim* ⇒ *Hero leaves on mission*
- Plausibility measures for elements of narrative
  Father is an accountant → *high*
  Father is a robber chief → *medium*
  Father is a plant pot → *low*

Many of these can be provided by statistical modelling and automatic extraction from text corpora. For example, the OpenIE project addresses acquisition of factual knowledge from web text. We've recently been focusing on the problem of extracting event knowledge from text.

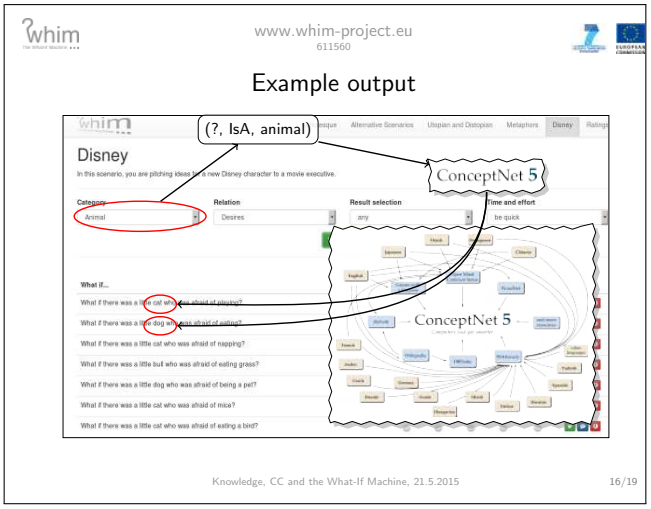
2.3 Progress



The project's currently halfway through. There's a prototype of all the components connected up, in a rather basic form. It has a web interface – go and try it out for youself!

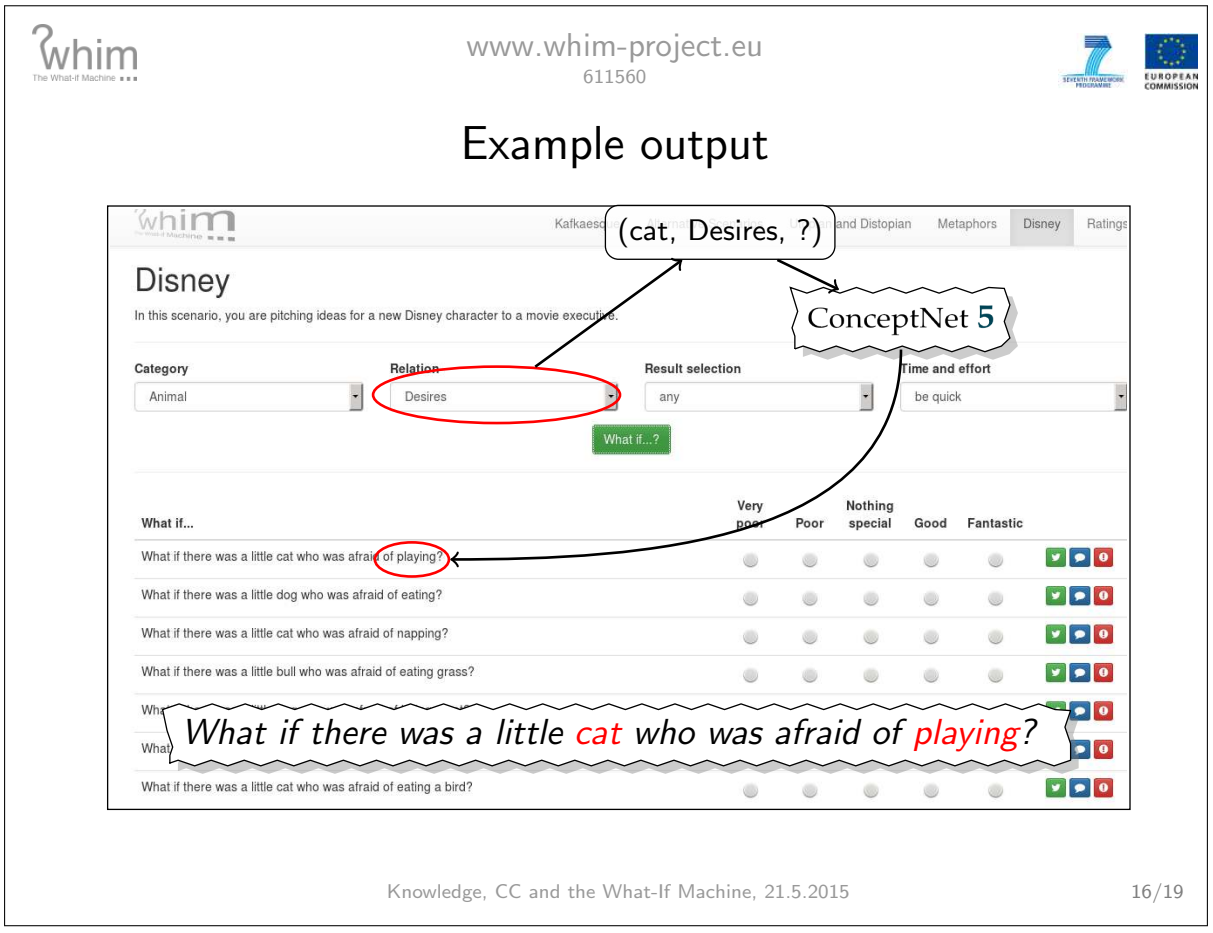
This initial system includes a variety of techniques to generate ideas. Some basic narrative generation systems are up and running, outputting some simple metrics. It includes a human feedback system, as you'll see on the web interface: this will be used for machine learning on narrative generation metrics.

One key thing that has come out of the experience of getting this prototype set up relates to the format in which information is passed around between the components of the system. Once what-ifs have been generated, it turns out that the *what if dogs had wings?* type of sentence doesn't include enough information about the process that generated it. The solution is what we're calling **mini-narratives**. These have a little more structured information that just the simple idea, but are not expanded into full narratives – that's the job of the narrative generation component.



Here's an example of one of the ways of producing what-ifs in the current system. It's in the web interface, so you can try it yourself if you want. This method uses ConceptNet – a large world knowledge resource, compiled from automatic information extraction, with human curation. We say we want an idea about animals and it uses ConceptNet to look up some animals. For each one, it looks up things they like doing. It then produces a what-if by some simple template-filling: *cat likes playing* ⇒ what if *cat was afraid of playing*?

This isn't a very interesting technique, but it's one of many (some of the others are more interesting). However, it is able to produce a load of mediocre-looking ideas and, occasionally, a really good one. At the moment, this is not going through any filtering to speak of using the narrative generation technique.



3 The future

The What-If Machine in the future

Current issues

- Improving narrative generation → requires:
  - reliable world knowledge
  - knowledge of unusual subjects
  - richer event-based knowledge
    event sequences, actions + results, goals
- Collecting human feedback
- Analysing feedback
- Predicting idea quality from narrative metrics

We're currently working on improving the narrative generation part of the loop: better knowledge acquisition, ways to use that knowledge, and more metrics that the system can base its judgements on. We need more reliable world knowledge and, often, about subjects that are not typical targets for information extraction systems. We also need richer event-based knowledge to help string together narratives.

We're considering different ways of collecting human feedback and how this can be used by the system, as well as to evaluate the whole process.

The What-If Machine in the future

What might WHIM tell us?

- Knowledge/reasoning requirements of fictional ideation
- Plausibility of narrative-based mechanism
- Creative generation vs. traditional AI
- Evaluation of creative systems

How might WHIM be useful?

- Working with interested practitioners:
  - Marketing people
  - Artists
  - Games designers...
- Narrative generation → story realisation

So, what do we hope WHIM can tell us? It should give us (more) new insights into the requirements (for a system or a human) for fictional ideation. We hope to be able to draw some conclusions about the plausibility of this narrative-based mechanism as a means of modelling human idea generation and assessment. We'll certainly come out with some further insight into how the demands of creative generation differ from traditional AI. And we hope to be able to offer some new ideas about evaluation of creative systems.

Of course, WHIM has its practical uses, which we will be exploring during the next year. We're already working with people from various professions to whom automatic ideation could be useful: marketing folks, artists, games developers, etc. And seeing as we're working on narrative generation as one component of the project, an obvious way to produce output from the system is in the form of stories, based on some of the narratives it tries out.

I hope that I've managed to convince you that Computational Creativity is a hot topic! There are exciting big questions and many interesting applications. I've discussed some of the challenges facing the field and its accomplishments so far. In particular, I've tried to focus on the role of computational systems in exploring theoretical models of creativity, and WHIM is one example of this, demonstrating one way we can make progress in the understand of human creativity through a computational approach.