Situated interpretation in computational creativity

Nick Kelly

Australian Digital Futures Institute, University of Southern Queensland, Toowoomba, Australia

John S Gero

Krasnow Institute for Advanced Study and Department of Computational Social Science, George Mason University, United States of America; and Department of Computer Science and School of Architecture, University of North Carolina at Charlotte, United States of America

Corresponding author: Dr Nick Kelly
Email: nick.kelly@usq.edu.au
Phone: +61 409 652 047
Address: Room A222, University of Southern Queensland, Sinnathamby Boulevard, Springfield, Queensland, Australia, 4066

Situated interpretation in computational creativity

This paper describes, formalizes and implements an approach to computational creativity based on situated interpretation. The paper introduces the notions of framing and reframing of conceptual spaces based on empirical studies as the driver for this research. It uses concepts from situated cognition, and situated interpretation in particular, to be the basis of a formal model of the movement between conceptual spaces. This model is implemented using rules within interacting neural networks. This implementation demonstrates behaviour similar to that observed in studies of human designers.

Keywords: interpretation; computational creativity; situated; design;

1. Introduction

Attempts to understand, support and automate aspects of human-like creativity are grounded in the notions of search and transformation of a space of possible solutions (Boden, 1991; Langley, 1987). Within this paradigm for computational creativity, a system may discover useful and novel or surprising artefacts (in the P-creativity sense), through search within a defined space or through exploration that transforms this space in some way (Boden, 1991; Gero, 1994; Wiggins, 2006b). Creative systems have been produced that can successfully search or transform an identifiable space to produce P-creative (and potentially H-creative) artefacts in diverse domains such as architecture (Merrell, Schkufza, & Koltun, 2010; Schneider, Fischer, & König, 2011), art (Colton, 2012; McCorduck, 1991), mathematics (Colton, Bundy, & Walsh, 2000; Lenat, 1976) and music (Pachet, 2012; Smith & Garnett, 2012). A challenge for creative systems that has not yet been adequately addressed is the *framing* of creative tasks, the production and development of the space within which creative activity occurs (Cross, 2004; Dorst & Cross, 2001; Schön, 1983; Seelig, 2012).

For systems aiming to frame creative activity in a way that is inspired by human phenomena the literature suggests that: (i) the system will have knowledge from experience;

(ii) the system will draw upon these experiences to set up the space within which creative activity will occur; and (iii) the system will change this space during creative activity. For example, in studies where designers 'think aloud' whilst designing it has been observed that designers are able to re-interpret their work in a novel way that changes their understanding of what it is that they are doing (Schon & Wiggins, 1992; Suwa, Gero, & Purcell, 2000; Suwa & Tversky, 1997). The designer has produced a design artefact within one framing of the problem – and then, from within this frame, been able to find entirely unexpected features within the same artefact.

In this paper a situated framework is articulated and implemented to explain the interaction between experience, expectation and a changing frame for a creative task. The process of interpretation within a creative system is where this interaction occurs, due to the clear distinction between the thing being perceived (e.g. an image of a pipe) and the interpretation of that thing (e.g. it need not be interpreted as a pipe). Each time a system interprets, we may ask the question why it produced this interpretation and not another. The claim being made is that for systems aiming at human-like creativity, movement between frames can be triggered by interpretation, and that this can be modelled and explained as the interaction between experience (what the system knows), expectation (what is in and implied by the current frame) and the stimulus (what is being interpreted).

Adapting nomenclature from Wiggins (Wiggins, 2006a) two different spaces can be identified for a system. The first is the *universe*, the space of artefacts potentially accessible to the system without limits upon time or resources. In many creative systems (e.g. any that permits an agglomerative production rule) the universe is an infinite space. Within a particular state of the system creative activity takes place in a smaller space within this universe, based upon the experiences (or knowledge) of the system and the notions to which it is currently attending. This reduced space will be referred to as the *conceptual space* of the system.

These two spaces are illustrated in Figure 1, inspired by studies of designers engaged in creative activity (Suwa et al., 2000; Suwa & Tversky, 1997). The rectangle in Figure 1 represents the universe of the designer. Within this space the designer searches for a solution within the limited conceptual space (grey ellipse), a space that is constrained by the designer’s conception of the design task as well as their past experiences. Something causes a change to the conceptual space, leading to a new space that can potentially be highly dislocated from the preceding space. This kind of a dislocated movement in conceptual space is sometimes described as a ‘moment of insight’ (Csikszentmihalyi & Sawyer, 1995).

This paper describes and models the way that the process of interpretation can move a system from one conceptual space to another in a way that is useful to the creative task. It occurs through the interaction between the conceptual space, the implicit expectations of that space and the stimulus being interpreted. The paper is structured by first introducing notions of situatedness and interpretation, followed by the formulation of simple examples of systems to distinguish situated interpretation, followed by an implementation of situated interpretation. The paper concludes with a discussion of the significance of this modelling.

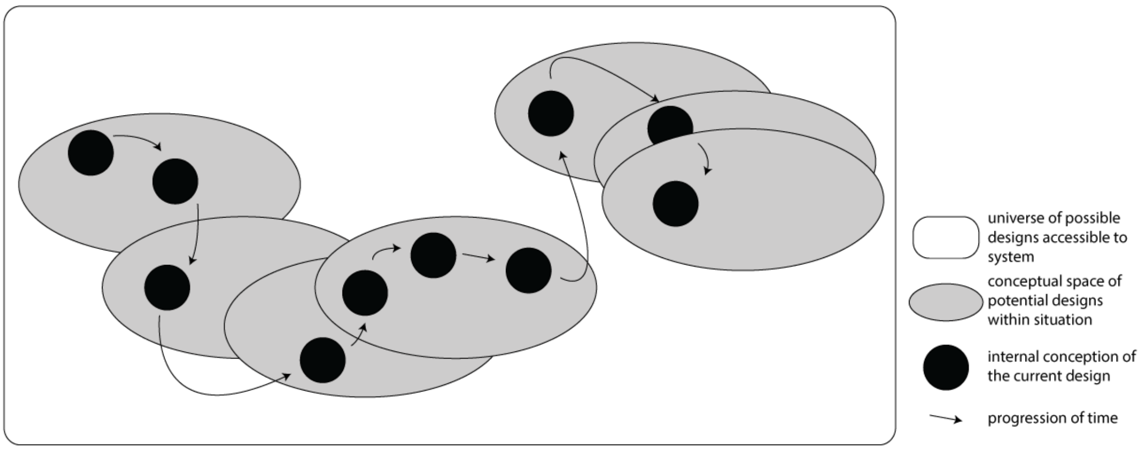


Figure 1 Movement between conceptual spaces during creative activity (after Kelly & Gero, 2014)

2. Theoretical background

2.1 Situatedness

In a *situated* system knowledge is something that is developed through experience of interaction with the world and is constrained by the way that the system conceives of its own activities (Clancey, 1997). As the system continues to experience the world, "subsequent experiences categorize and hence give meaning to what was experienced before" (Clancey, 1997; Dewey, 1896). An example of this can be seen in the way that perceptual symbol systems (PSS) represent and utilize concepts (Barsalou, 1999). Concepts in a PSS are conceived as convergence zones that co-ordinate the re-enaction of elements of, rather than whole entities of, perceptual experiences. This re-enacting occurs within and is a function of the current conceptual space, in contradistinction to the notion of concepts as static identifiers that are stored and retrieved (Barsalou, 2005a, 2005b). An implication of situated enaction of perceptual experiences rather than retrieval of static concepts (a higher level of abstraction) is that the combinatorial possibilities from those perceptual experiences are exponentially greater.

In this work situations are considered as a construct emerging from experience with the co-ordination of concepts. A situated system is one in which the co-ordination of concepts changes. Similar definitions that assist in clarifying what is meant by this are those systems in which the internal context changes (Kennedy & Shapiro, 2004), the epistemic frame changes (Shaffer et al., 2009), the ecology of mind changes (Gabora, Rosch, & Aerts, 2008) or the use of grounded knowledge from experience within the world changes (Barsalou, 2007).

2.2 Situated interpretation

Interpretation is defined as a process by which the experiences of the system are used to create an internal representation from a source, where the term source refers to the artefact (internal or external) being interpreted. *Situated interpretation* is said to occur in systems that: (i) interpret; (ii) are situated; and (iii) utilise expectations in interpreting. It is a process through which a source, the current conceptual space and the past experiences of the system interact to produce an internal representation. Change to the conceptual space can occur during this process.

2.2.1 Expectations in situated interpretation

One type of interpretation can be seen in systems that relate a source to one of a collection of static identifiers through a relationship of 'as a', e.g. identifying an unknown object as a BLOCK (Pylyshyn, 1977; Russell, Norvig, Canny, Malik, & Edwards, 1995). In contrast to this, a situated interpretation system commences with an expectation of what will be interpreted, and proceeds to construct an interpretation based upon a 'pull' from these expectations and a 'push' from the source to produce an internal representation (Gero & Kannengiesser, 2004; Kelly & Gero, 2014; Kelly & Gero, 2011). In an unchanging or constrained environment a system may be able to develop expectations that are useful for all circumstances. However, in a dynamic or unbounded environment a system will likely find circumstances in which adaptation of expectations is required. Interpretation is concerned with this need for a balance between a push from the "buzzing blooming confusion" of a source (James, 1890) and a pull from the stability of expectations.

Through pull, interpretation attempts to construct an internal representation of the world that fits with what is expected. The expectation is present prior to the stimulus, with pull attempting to see whether it can adequately construct what is expected using the data present in the stimulus. For example, when participants in an experiment are played the sound of a single note followed by the sound of white noise they are able to 'hear' the note within the white noise (Riecke, van Opstal, Goebel R, & Formisano, 2007). The expectation of the note prior to the white noise forms the basis for perceiving a note within the random signal.

Push is the part of interpretation concerned with data that are not expected that may still require perception and allows for expectations to change based upon what is found in the source. Push deals with those circumstances where for a number of reasons expectations might not be useful (e.g. not a good fit with the world). An example of push from the source into interpretation is the way that the sound of a police siren is heard even if it is not expected.

The model of interpretation presented and implemented here contains both pull and push. In order to implement pull a notion of expectations is required. A distinction can be made between explicit and implicit expectations. Explicit expectations refer to the concepts and percepts currently being attended to within a conceptual space. Implicit expectations are then those that are related in some way to these explicit expectations (e.g. through similarity or through proximity within knowledge structures) but not currently being attended to. Examples of how implicit expectations affect interpretation can be observed in human phenomena of implicit memory (Graf & Schacter, 1985; Schacter, 1987), priming (Schacter, 1987) (Phaller & Squire, 2009) and selective attention (Cozolino & Siegel, 2009).

2.2.2 Examples of situated interpretation

Aspects of situated interpretation can be identified in existing knowledge-based and creative systems. Formal models of learned selective attention bias the production of an internal representation on the basis of past experiences (Kruschke, 2011). There are many examples of analogy making systems (Goel, 1997; Goel, Vattam, Wiltgen, & Helms, 2012; Jeong & Kim, 2014) that are concerned with mapping relationships in function, behavior and structure from a source to a target (Gentner, 1983; Gentner & Colhoun, 2010; Qian & Gero, 1996).

This ability to find analogical similarities is an important part of the ‘pull’ aspect of interpretation, as the construction of an internal representation based upon expectations will be enhanced by analogical reasoning; such reasoning is also often an example of human creative activity (Green, Kraemer, Fugelsang, Gray, & Dunbar, 2012; Holyoak, 1996). Models of conceptual slippage can be considered one way of implementing the push-pull of interpretation (French, 1995; Hofstadter, 2008).

2.3 Towards a formal description of situated creative systems

The elements of Wiggins’ formal model of creativity (Wiggins, 2006a, 2006b) provide the basis for formulating a model of a situated creative system, Table 1. The aim of developing this model is to consider ways that a situated system can navigate its own knowledge to produce P-creative artefacts in a manner inspired by human studies (Schön, 1983; Suwa et al., 2000).

Table 1 Symbols used in formulating situated computational creativity

$[\![.]\!]$	A function generator for forming conceptual space
$\langle\langle .,.,. \rangle\rangle$	A function generator that moves the system from one situation to another
C	Conceptual space
E	Expectations
I	Situated interpretation
L	The language of the system
N	A subset of L for interpreting during T
R	A subset of L used in generating the conceptual space C
S	Situation
T	A subset of L defining traversal within C
U	The universe of conceivable concepts within the system

2.3.1 The universe and conceptual space

The universe U is defined as the concepts that the system is capable of producing given a language, L. The elements of the universe that have been experienced by the system will be referred to as $U_!$.

A function $[\![.]\!]$ generates a conceptual space C from the experiences of the system $U_!$ and a set of rules R, Equation 1. It is an abductive task to produce this function.

$$C = [\![R]\!](U_!) \quad (1)$$

2.3.2 Exploring and interpreting within conceptual space

The system carries out some activity defined by a set of rules T that enables search to occur given a particular conceptual space, C.

As a part of this search of conceptual space the system produces representations. These representations are interpreted to create an internal representation using a set of rules, N.

2.3.3 Movement between situations

An interpreter generates functions using the three arguments R, T and N that result in the system moving from one conceptual space to another, through changes to R, Equation 2 (Wiggins, 2006a).

$$C_{i+1} = \langle\langle R, T, N \rangle\rangle(C_i) \quad (2)$$

In a situated system, previous experiences are used in constructing the conceptual space. Traversal T within the conceptual space is dependent solely upon a subset of previous experiences that are re-enacted within the situation, not upon all previous experiences $U_!$.

3. Distinguishing situated interpretation in simple creative systems

A simple creative system serves to clarify the notion of situated interpretation for movement between conceptual spaces. An abstract description is given followed by two different instantiations. In this system the language L permits all real numbers $\mathbb{R}$ and U is an infinite space. The system has had experience of a subset of this universe, $U_!$, which is limited to the integers $\{1, 2, 3, \ldots, 10\}$. R is used to generate the conceptual space C from experience by attending to two concepts within $U_!$ such that:

$$[\![R]\!](U_!) = C = \{c_1, c_2\}$$

Creative activity occurs as expressed in Equation 2. Two different systems will be described, one in which the system changes situation through T and the other through I. The latter is representative of situated interpretation.

3.1 A simple situated generative system

The system is situated, and T is defined for this system as a function that produces an external representation of a design, x, using the average of the current two concepts within the conceptual space. The reason that just two concepts are used is that it is the smallest number required to show interaction; a larger number could also be used. The average is used here for generation as a simple way to represent an interaction between concepts. T is productive in that the system can produce things it has never experienced:

$$T(C) = average(c_1, c_2) = x$$

Interpretation within the system is unaffected by expectations (i.e. interpretation is not situated). An internal representation c_x is created of the source that the sensor encounters, and for the sake of the example perfect sensors are assumed in the system along with an ability to learn such that:

$$N(x) = c_x = x$$

The conceptual space is changed using R to select two concepts and the process of creative activity can repeat in this way. Such a system is capable of having an unlimited number of experiences.

3.2 A simple situated interpretation system

A similar system can demonstrate a similar unlimited number of potential experiences through situated interpretation rather than through generation. In this system N is productive and T is not. T involves selection of a concept from experiences and use in generating external x:

$$T(U_!) = x$$

The conceptual space of the system consists of the two currently attended to concepts c_1 and c_2. In order to demonstrate interpretation as a construction from expectations the system requires some form of explicit expectations. This is implemented in this model as an expectation to find the average value of the concepts in the situation. In this way expectations are situated – the same $U_!$ with a different C will produce different expectations. Interpretation N in this example is represented in a simple way, as taking an average between the expectation and the source. This is a simplistic representation of the negotiation between the pull from the expected value and the push from the value being sensed:

$$N(x) = c_x = avg(avg(c_1\,, c_2), x)$$

The interpreted value resulting from N is now an experience, a part of $U_!$. The process of creative activity repeats in this way. Such a system is capable of having unlimited novel experiences.

3.3 Discussion: search in infinite space

To an external observer both systems exhibit a similar search of the space. The system described in Section 3.1.2 is achieving what could be described as search within an infinite space through interpretation in a constructive way rather than through generation. This is a simple example to pose the question of how constructive interpretation can lead a system with previous experiences towards spaces within which useful concepts might be encountered – systems that through creative activity develop a useful framing of the problem (Maher, 2000).

The system in Section 3.1.2 can be conceived as a system that searches space. By changing its conceptual space it is capable of discovering any one of the uncountable concepts $c \in \mathbb{R}: 1 < c < 10$. Table 2 represents the system using pseudo-code to make this clearer. The purpose of the example is to distinguish interpretation from generation and show its effects. The distinction of search-and-exploration through interpretation rather than generation will become clearer through subsequent examples with deeper knowledge structures.

Table 2 Pseudo-code as search for real numbers

$Init$	Select a target as any real number $\alpha \in \mathbb{R}: 1 < \alpha < 10$ $U_{!} = \{1,2,3, \dots ,10\}$ Produce an external concept x as initial source
1	R – Select from all known concepts the closest upper and lower bound over the target
2	N – Interpret x to produce new concept c_x
3	T – Generate new x from within C
4	Repeat from step 1

In the system described in Table 2 the situation gets progressively closer to one in which the targeted concept can be located. $U_{!}$ expands as the system has more experiences. The purpose of the separation between generation and interpretation is to provide a contrast –

there is no reason why both cannot be implemented within the one system. The claim is that a system with significant experience can undertake creative activity $\langle\langle R, T, N\rangle\rangle$ within a conceptual space that is a reduced subset of experience $C \subset U_!$ that results in a move towards a space C that is useful to the current task. It does this through interpretation that relates expectations (experience within the current situation) to the product of design activity x, be it internal (e.g. thought experiments) or external (e.g. a sketch).

4 Situated interpretation in more complex knowledge systems

4.1 Situated interpretation in knowledge based systems

How can situated interpretation be useful in knowledge-based creative systems? Knowledge in such systems is assumed to take the form of experiences from multi-modal sensory data and potentially complex internal knowledge structures, typically a hierarchy. Through taking actions in their world and through sensory observation these systems develop knowledge about the sense data produced by the world, with potential for abstraction over experiences (Barsalou, 2005a).

Such a system can hold expectations about the world based upon its knowledge – of spatial/structural relations (e.g. seeing a car implies four wheels, whether or not they are visible) and temporal relations (e.g. an object will continue at its present velocity). The focus in this description is limited to spatial expectations, as this is sufficient to describe and implement situated interpretation.

Let expectations E be made up of explicit expectations E^* (those to which the system is attending) and implicit expectations E' (those implied by the structure of the system and resulting from experience).

In order to introduce a hierarchy of abstract we will refer to a concept $c \in C$ as a prototype resulting from convergence of perceptual information. Units of perceptual information will be referred to as percepts $p \in P$, which themselves originate in sensors (Gärdenfors, 2000). The situation s is an explicit representation of the co-ordination of concepts. Within a creative system explicit expectations can be defined as those concepts in the conceptual space and the percepts from which they were constructed, Equation 4.

$$E^* = \{P, C, s\} \tag{4}$$

Situated interpretation has been defined as the construction of an internal representation from a source x and expectations E. Let the interpreted representation I be composed of different levels of abstraction:

$$I = \{I_P, I_C, I_s\}$$

A double arrow $\rightrightarrows$ is used which contains (i) a loop representing 'pull' or construction from expectations; and (ii) a solid arrow representing a 'push' from data where expectations are not satisfied. Interpretation occurs through three processes:

$$I_P: x \overset{E_P}{\rightrightarrows} I_P$$

$$I_C: P_I \overset{E_C}{\rightrightarrows} I_C$$

$$I_s: C_I \overset{E_s}{\rightrightarrows} I_s$$

Each of these represents a process by which the system attempts to construct an interpretation (i.e. checking to see if expectations are useful) and, if not, then the data is passed on to a higher layer (after Gero & Fujii, 2000; Hawkins, 2005).

Chaining these processes results in a conception of interpretation as a dynamic parallel process that is likely to produce coherency in I_P, I_C and I_s due to the feedback of expectations and feedforward of data, Figure 2. For example, pull in both P and C may be unable to construct from expectations and so in both cases data are pushed upwards. This leads to a change of expectations in s which then cascade down to C and then P. These expectations can now be used to produce I_P, I_C and I_s.

Interpretation in this way will typically settle upon a unique I over time. In rare cases the process may be unstable, oscillating between different I. This is a good fit with observations of human interpretation of optical illusions (e.g. (Leopold, Wilke, Maier, & Logothetis, 2002; Long & Toppino, 2004)). The model is inspired by the neuroscience observations of Mountcastle and the detection of feedback between layers of the human cortex (Hawkins, 2005; Mountcastle, 1997).

Pull from expectations maintains stability within current C. Push from the data in each layer when expectations are not met facilitate learning from new experiences and a change in C.

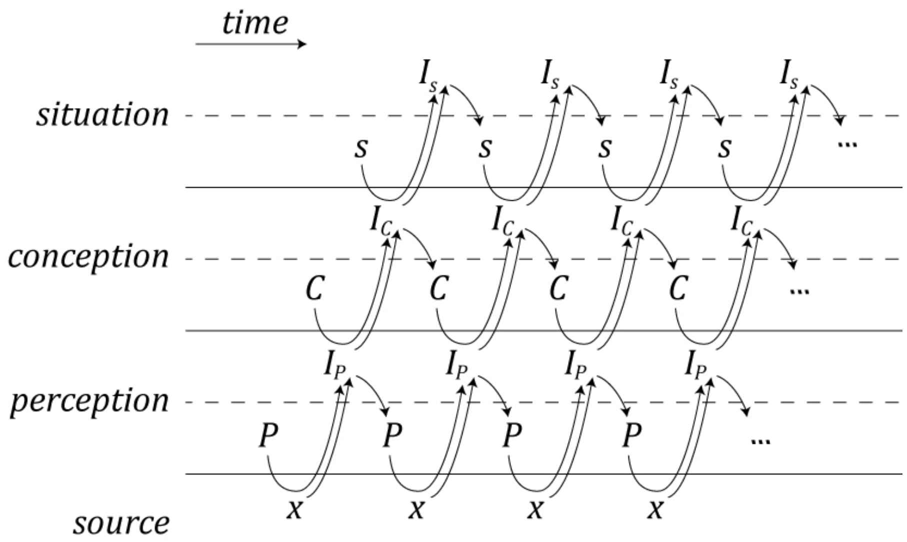


Figure 2 Situated interpretation as dynamic construction from a source using expectations

4.1.2 Implicit expectations in interpretation

This description is useful as a model for how these notions of cognitive processes (expectations) and sensory perceptions (of a source) can interact during interpretation. Whilst the focus of interpretation is upon construction of an internal representation of a source, a result of potential changes to expectations during the process is that C can change through interpretation. The significance for creativity is that such a form of interpretation allows for specific exploration through T within restricted C, whilst permitting change to conceptual space through N. It also addresses the question of how C changes when it does change – the experience and knowledge structures of the system are, in the context of C, the implicit expectations and it is these that suggest how the system will move to a new C.

The set of implicit expectations E' is a function of both the past experiences $U_!$ and the current situation E^*. The function f for producing implicit expectations is implemented using some notion of proximity to E^*:

$f(U_!, E^*) = E', E^* \cap E' = \emptyset$

The notion of proximity is dependent upon the way that a system represents knowledge. In the systems in Sections 4.2 and 4.3, proximity is implemented in two ways, as a function of Euclidean distance within a vector space and as connectedness between layers of abstraction. Whilst the specific measure of proximity is implementation-specific, some notion of proximity is required; this may be based upon notions such as similarity (Nosofsky, 1988) or upon something entirely different, e.g. proximity within a computational representation.

Secondly, the concepts held by the system have arisen through experience. Consider identically implemented systems A and B, each of which have had different experiences such that $U_!^A \neq U_!^B$. It is entirely feasible based upon common experiences, $U_!^A \cap U_!^B$, that explicit expectations will be identical, $E^{*A} = E^{*B}$, yet due to perhaps minor differences in past experiences the implicit expectations will be different, $E'^A \neq E'^B$. This can lead to significant differences between interpreted representations $I(x)$ in each system even if the difference in experience is trivial. It is possible (if unlikely) for systems to have different experiences yet identical explicit and implicit expectations. Table 3 summarizes these relationships between past experiences, implicit and explicit expectations and interpreted representations.

During creative activity, interpretation makes use of both the explicit and implicit expectations. It is the presence of implicit expectations that forms the basis for a claim that through situated interpretation changes to C occur based upon past experiences that can aid a system to move to a C within which a desired solution can be found.

Table 3 Relationships between experiences, implicit and explicit expectations and interpreted representations of a source in two systems A and B

Experiences	Explicit expectations	Implicit expectations	$\exists I: I^A \neq I^B$	$\exists I: I^A = I^B$
$U_!^A = U_!^B$	$E^{*A} = E^{*B}$	$E'^A = E'^B$	No	Yes
$U_!^A = U_!^B$	$E^{*A} = E^{*B}$	$E'^A \neq E'^B$	Yes	Yes
$U_!^A = U_!^B$	$E^{*A} \neq E^{*B}$	$E'^A = E'^B$	Yes	Yes
$U_!^A = U_!^B$	$E^{*A} \neq E^{*B}$	$E'^A \neq E'^B$	Yes	No
$U_!^A \neq U_!^B$	$E^{*A} = E^{*B}$	$E'^A = E'^B$	Yes	Yes
$U_!^A \neq U_!^B$	$E^{*A} = E^{*B}$	$E'^A \neq E'^B$	Yes	Yes
$U_!^A \neq U_!^B$	$E^{*A} \neq E^{*B}$	$E'^A = E'^B$	Yes	Yes
$U_!^A \neq U_!^B$	$E^{*A} \neq E^{*B}$	$E'^A \neq E'^B$	Yes	No

4.2 Implementing situated interpretation

4.2.1 Abstract framework for implementation in unsupervised learning systems

A general description of implementing situated interpretation is provided for two linked unsupervised learning systems in which inputs are classified by finding the best matching node BMU (e.g. SOM or ART networks (Ciresan, Meier, & Schmidhuber, 2012; Gu, 2010)). A system with 2 layers, L_1 (lower) and L_2 (upper) is used and each layer is described as a collection of nodes M.

The output from L_1 forms the input to L_2 as in hierarchical unsupervised learning systems. The algorithm for finding the BMU for an input is altered such that only a subset of nodes, the expected nodes $M^E \subset M$, is utilised. M^E is constructed from the explicit expectations M^* within each layer and implicit expectations M'. Explicit expectations M^* are those nodes currently active in L_1 and L_2.

The set of implicit expectations are constructed from a combination of: (i) distance within a layer (either based upon node adjacency or distance of nodes from those in M^*); and (ii) connections between layers (from experience).

A similarity threshold σ is defined as a minimum similarity required for construction from expectations to occur within a layer utilising solely the expected nodes. and is expressed as a distance d (e.g. based upon Nosofsky similarity (Nosofsky, 1988) between input and BMU). Construction from expectations can occur within a layer, e.g. from a source x, if the inequality in Equation 3 holds:

$$\{ BMU \in M^E | d(BMU, x) < \sigma\} \tag{3}$$

If construction cannot occur within a layer then data are instead pushed to an upper layer, where the same process occurs. In this way either: (i) expectations are satisfied in an upper layer, leading to changed expectations in a lower layer; or (ii) the topmost layer does not have expectations satisfied.

Whilst constructive interpretation is described as a dynamic process, Figure 2, it can be implemented as a linear process within a two-layer network, Figure 3. As a linear process there is first an attempt by pull within each layer to construct from expectations, with push occurring where this construction does not occur. This is repeated in each layer, cascading upwards; in Figure 3 it occurs first in L_1 with over external data, and then in L_2 over the current output from L_1. If expectations change in L_2 then these are fed back down to L_1.

When no interpretation has been constructed in any layers through M^E one of three things happen: (i) learning occurs in a layer; (ii) a change can be triggered in σ; or, as in the implementations here (iii) in the top layer implicit expectations are expanded through some form of spreading activation (Anderson, 1983).

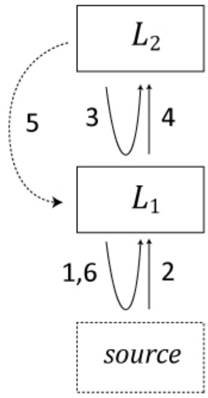


Figure 3 A linear implementation of situated interpretation in a two layer system with numbered steps: 1 pull from expectations in L_1; 2 push from the source into L_1; 3 pull from the expectations of L_2; 4 push from L_1 into L_2; 5 an update of L_1 expectations by L_2; and 6 interpretation through pull from L_1 with the updated expectations.

4.2.2 An implementation demonstrating stability

The architecture described in Section 4.2.1 was implemented using two linked 2D self-organising maps (SOMs) (Dittenbach, Merkl, & Rauber, 2000; Kohonen, 1990), SOM_1 and SOM_2, with sizes respectively 100x100 nodes and 40x40 nodes. The model demonstrates the pattern of changing C observed in Figure 1. Three outputs from the lower layer form input to the upper layer, which during training leads to abstraction over these inputs, Figure 4. During training, each SOM uses the Kohonen training algorithm (Kohonen, 1990) in two phase training to: (i) reducing the neighbourhood radius to 1; and (ii) reducing the learning increment from 0.1 to 0.

The network is trained upon sets of three 16x16 icons. Each set of three is constructed from randomly generated crosses, squares and triangles – both size of the icon and placement within the 16x16 tartan grid are random. Further, every set of three icons during training is composed of either: (i) all three icons the same type (three squares, three crosses or three triangles); or (ii) all three different (e.g. one cross, one square, one triangle).

Following training, an initialised system holds a set of explicit expectations that guide interpretation and the system is placed in a changing environment. The environment is automatically generated, with a state generated using the same algorithm that produced the training set. The environment moves towards another one of these expected states through a number of interpolation states that are generated through the addition or subtraction of pixels, until the next expected state is reached. This process continues indefinitely. The left side of Figure 5 demonstrates this progression from one expected state to another with two interpolation states in between. The right side of Figure 5 shows the resulting internal representation.

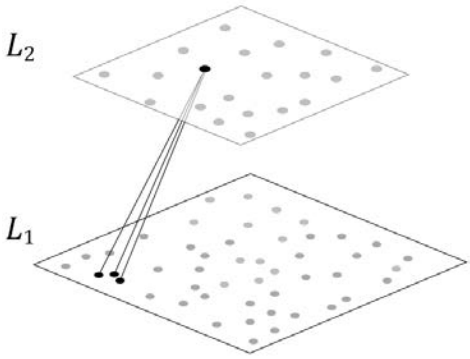


Figure 4 Explicit expectations in a two-layer SOM

The resulting behaviour in the system is stability of C with repeated interpretation of a changing x; followed by a sufficiently different x to trigger a change in C. In Figure 5(a) the system interprets the source as three crosses. The source x is similar to the training stimuli and, with current E, Equation 3 is satisfied, producing $N(x) \approx x$. In Figure 5(b) $N(x)$ is unchanged, despite there has being a change in x. This occurs because Equation 3 is still satisfied for current E, and thus the internal representation does not change.

In Figure 5(c) Equation 3 is no longer satisfied due to further change in x increasing d. A change in $N(x)$ has happened producing interpretation as a cross, a triangle and a square. This change occurs as described in Figure 3, with L_1 unable to satisfy Equation 3, passing data to L_2 which in turn cannot satisfy Equation 3. Spreading activation in L_2 leads an interpretation in this layer that can satisfy Equation 3. This causes a change of E in L_1. With these new expectations an interpretation is found in L_1 that fits with L_2. The changed interpretation found in L_2 is produced through the implementation of implicit expectations as based upon Euclidean distance within the layer. For this reason there is similarity in I_1 between Figure 5(b) and 5(c) – the new interpretation is based upon the implicit expectations of the previous situation and the nature of x.

In Figure 5(d) x is now four squares. This x is one of the expected states, data from the training set and so a part of $U_!$. However, due to the situation at this time the system maintains an unchanged $I(x)$ as the data satisfies Equation 3 with current E. The interpreted representation is the one that fits where the system is at (internally) at the time of interpreting, which will often not be the 'best match'.

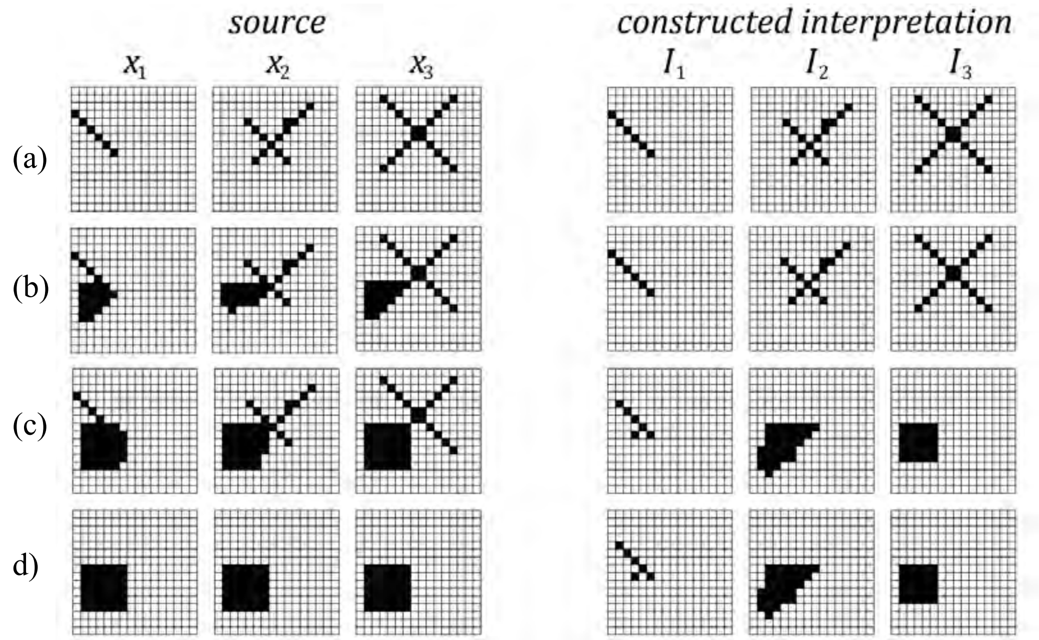


Figure 5 Constructive interpretations in a changing environment, left side as the environment and right side as constructed interpretation from: (a) three crosses; (b) first interpolation state; (c) second interpolation; and (d) four squares

4.3 Divergent exploration with floor plans through situated interpretation

A further implementation serves to demonstrate situated interpretation within creative systems, in the domain of housing floor plan designs. The system is not a creative system by most definitions as it has neither goals nor a fitness function. However, it demonstrates further the balance between specific and divergent exploration important for creative activity (Greif, 1994; McCrae, 1987). The system is initialised within a conceptual space and repeats a cycle of design-generation, interpretation and then generating once more within the resulting conceptual space – an implementation the “seeing-moving-seeing” of design (Schön, 1983), Equation 2.

A set of 56 floor plans from 3 different architects (Palladio, Wright and Khan) was used for training the system, Figure 6. In training the system, automated edge detection was used to separate each plan out into a number of 16x16 feature maps, with these feature maps the input into L_1. Training in L_2 has supervised connections between hard-coded floor plan representation in L_2 (rather than learnt through training) and feature maps in L_1.

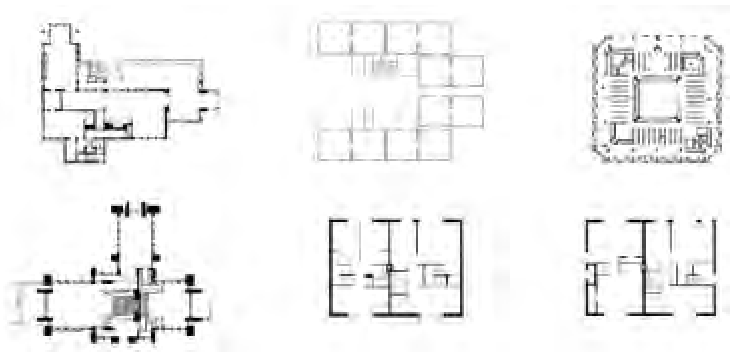


Figure 6 A sample of 6 of the 56 floor plans used in training the system (after Jupp, 2005)

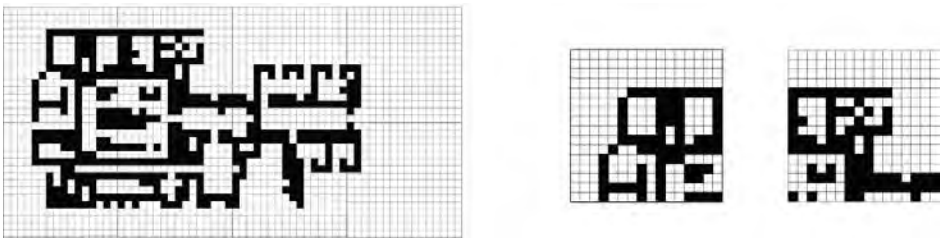
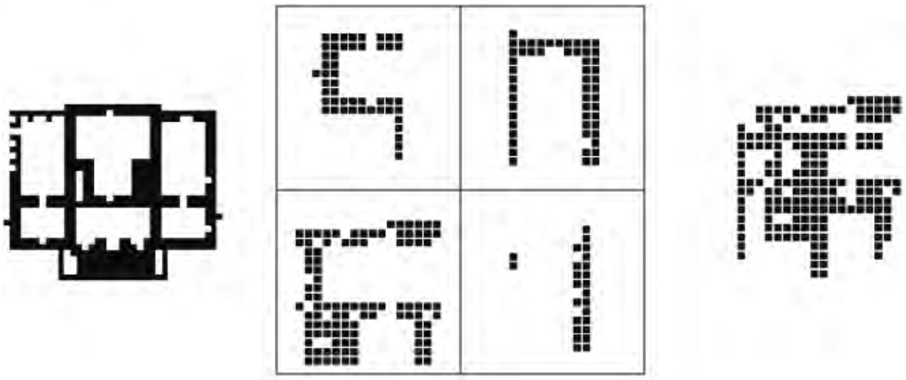


Figure 7 The floor plans with full detail (Figure 6) are reduced to a canonical representation and learnt as a sequence of 16x16 pixel icons

4.3.1 Generation within the system

The system is initialized within a situation, represented by a single node within the top layer and C is composed of 4 currently active concepts randomly selected from within the situation.

The system implements T as generation of an artefact in a 24x24 pixel canvas. The four currently active concepts are each placed within the canvas in a distinct location. For reference Figure 8(a) shows the initial representation of a floor plan, along with Figure 8(b), the four initial concepts of the system, and (c) a ‘sketch’ that has been produced by representing each of these concepts in a random location within the external canvas.



(a) (b) (c)

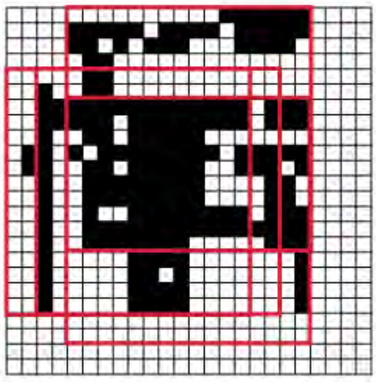
Figure 8 (a) A representation of the original floor plan; (b) a set of four feature maps as the current conceptual space; and (c) the representation produced by generating from within this space

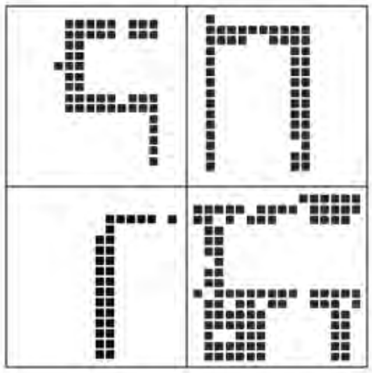
4.3.2 Interpretation within the system

After producing a representation, the system interprets it. The system saccades across the design from top left to bottom right and each 16x16 pixel section of the canvas encountered in this way is interpreted.

In Figure 9(a) the system interprets the canvas within its current situation. The red boxes indicate where, in the course of a saccade across the canvas, the system has been able to construct from expectations. It is typical of the system that it draws with concepts and then finds the same concepts within its own work. However, at some point interpretation will construct a concept from implicit expectations.

Figure 9(b) shows the concepts that have been constructed during interpretation. The top two and lower right images are all explicitly expected concepts. However, the lower left image is from a different floor plan – it has come from expectations implicit within the situation. It now becomes an explicit concept, used for sketching. In this way the cycle of moving and seeing continues.





(a) (b)

Figure 9 Interpretation within the system occurs through a saccade from top left to bottom right: (a) the four areas used for construction during interpretation marked by red boxes; (b) the concepts constructed through interpretation, where the lower left comes from implicit expectations.

4.3.3 Implicit expectations in the model

Implicit expectations in the model were defined as: (i) those nodes within L_1 within a specified distance from explicitly expected nodes; and (ii) those nodes within L_1connected to the current situation through connections in L_2. An example of the latter can be seen in Figure 9(b), where the implicit expectation used (lower left concept) has come from another floor plan that was not the initial situation. This other floor plan has other associated concepts which now become implicitly expected.

Without implicit expectations, the model will behave in a predictable way – explicit expectations will be used for constructing every time. As a result the interpretation process here has been programmed to prioritise the construction from an implicit expectation over an explicit expectation.

4.3.4 Discussion and limitations of the model

The model takes the constituent parts of an original floor plan and uses them to generate designs. The point of interest is that implicit expectations within the situation, through similarity in L_1 and connectedness through L_2, lead the model to change its conceptual space.

With these changes to C, as in the model in Section 4.2, a consistent state is maintained that allows for specific exploration, followed by identification of features during interpretation that shifts the system to divergent exploration. This fits with the pattern observed in designers in previous studies (Figure 1 and (Suwa et al., 2000)) as well as calls for both specific and divergent types of exploration to be present in models of computational creativity (Greif, 1994).

Further, when a change of C is triggered, the way in which the system moves into a new situation is determined by a combination of the current situation, past experiences and the implementation of the system. As observed in Figure 5(c) the new interpretation has something in common with both the previous interpretation and the new source. It is claimed that this type of interpretation can lead to the type of design trajectories observed in human creativity.

The systems implemented here are limited in a number of ways. Firstly, the model was implemented to maintain 4 concepts as the output from every process of interpretation. This was done by relaxing the value for σ during interpretation until at least 4 concepts were encountered (and if more than 4 concepts were found, priority was given based upon greater similarity). This limitation relates to the far broader question of 'chunking' in interpretation, which can be framed as the problem of how a system, interpreting a complex source, structures the source into a number of parts, each of which are interpreted. Situated interpretation as seen in Figure 2 taken with examples in the literature (Barsalou, 1999;

Gärdenfors, 2000; Hawkins, 2005) provide some clues as to how this may occur, yet this problem lies beyond the scope of the paper.

The most significant limitation is that the models implemented here use a simple technology (linked SOMs) and have just two layers. This has been done for the sake of clarity, yet a consequence is that the application of the technique to a real-world domain is yet to be demonstrated and evaluated. Further research is required to consider what effect the dispersion of implicit expectations between many layers of abstraction might have. Specifically, deep learning neural networks (Bengio, 2009) have been identified as a novel technique that would be well suited to adaptation for situated interpretation, with multiple layers of abstraction and high performance in visual and auditory domains which are well suited to demonstrations of computational creativity (Boulanger-Lewandowski, Bengio, & Vincent, 2012; Liapis, Martınez, Togelius, & Yannakakis, 2013).

5. Conclusion

Systems which aim at understanding and supporting human creativity can benefit from implementing situated interpretation. Situated interpretation is a novel paradigm for interpretation that has arisen from the situated cognition tradition. It is important to computational creativity because it provides a way of addressing the framing problem.

Within a situated system that has a great deal of experience of the world there are many possible conceptual spaces within which it may undertake creative activity. This precondition is important, as for systems with a great deal of experience, especially those with a type of situated conceptualization, there is rapid combinatorial explosion if the system attempts to create using the entirety of its experiences.

Given these many different conceptual spaces, how does the system move from the one within which it commences the problem, towards one which is useful for finding a solution? The answer suggested by situated interpretation is that this occurs through a form of exploration within the current situation (e.g. reasoning or action such as generation of an artefact) followed by interpretation. The expectations, both implicit and explicit, in current situation are used during interpretation such that as a result of interpretation the system may either: (i) maintain the current situation; or (ii) move to a new situation.

Where this movement to a new situation occurs, it is due to implicit expectations within the situation. In systems with multiple levels of abstraction this may occur in any level of abstraction, with propagation of expectations (as well as interpretations) between layers maintaining consistency of internal representations.

The combination of these effects results in systems that engage in the type of movement between conceptual spaces observed in Figure 1, where both specific and divergent exploration are present, and where changes of expectations at higher levels of abstraction can trigger large changes in the situation, and changes of expectations at lower levels can trigger small changes to the situation.

It is accepted that humans engaged in creative activity often change the frame within which they are acting. It follows that creative systems seeking to emulate this form of creative process need a way of moving between different conceptual spaces. This paper has proposed an argument for why interpretation is the way in this can occur, with past experiences, the current conceptual space and the current stimuli all being utilised to create an internal representation, a process that can also move the system to another conceptual space. The models described here provide a foundation towards the formalization of framing within creative systems.

Acknowledgements

This work is supported by the US National Science Foundation under Grant No. CMMI-1400466

References

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior, 22*(3), 261-295.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and brain sciences, 22*, 577-660.

Barsalou, L. W. (2005a). Abstraction as dynamic interpretation in perceptual symbol systems. In L. Gershkoff-Stowe & D. Rakison (Eds.), *Building object categories* (pp. 389-431). Mahwah, NJ: Erlbaum.

Barsalou, L. W. (2005b). Situated conceptualization. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 619-650): Elsevier.

Barsalou, L. W. (2007). Grounded cognition. *Annual Review of Psychology, 59*(1), 617-645. doi: 10.1146/annurev.psych.59.103006.093639

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations and Trends in Machine Learning, 2*(1), 1-127.

Boden, M. A. (1991). *The Creative Mind*. London, UK: Weidenfeld and Nicolson.

Boulanger-Lewandowski, N., Bengio, Y., & Vincent, P. (2012). Modeling Temporal Dependencies in High-Dimensional Sequences: Application to Polyphonic Music Generation and Transcription. In J. Langford & J. Pineau (Eds.), *Proceedings of the 29th International Conference on Machine Learning (ICML-12)* (pp. 1159-1166): Omnipress.

Ciresan, D., Meier, U., & Schmidhuber, J. (2012). Multi-column deep neural networks for image classification *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2012* (pp. 3642-3649): IEEE.

Clancey, W. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*: Cambridge University Press.

Colton, S. (2012). The painting fool: Stories from building an automated painter. In J. McCormack & M. d'Inverno (Eds.), *Computers and creativity* (pp. 3-38): Springer.

Colton, S., Bundy, A., & Walsh, T. (2000). On the notion of interestingness in automated mathematical discovery. *International Journal of Human-Computer Studies, 53*(3), 351-375.

Cozolino, L., & Siegel, D. (2009). Contributions of the psychological sciences. In B. Sadock, V. Sadock & P. Ruiz (Eds.), *Kaplan and Sadock's comprehensive textbook of psychiatry 9th edition* (pp. 620-622). Baltimore: Lippincott Williams and Wilkins.

Cross, N. (2004). Expertise in design: an overview. *Design Studies, 25*(5), 427-441.

Csikszentmihalyi, M., & Sawyer, K. (1995). Creative insight: The social dimension of a solitary moment. In R. J. Sternberg & J. E. Davidson (Eds.), *The Nature of Insight*. Cambridge, MA: MIT Press.

Dewey, J. (1896). The reflex arc concept in psychology. *Psychological review, 3*(4), 357.

Dittenbach, M., Merkl, D., & Rauber, A. (2000). The growing hierarchical self-organizing map. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN)* (Vol. 6, pp. 15-19). Como, Italy: IEEE Computer Society.

Dorst, K., & Cross, N. (2001). Creativity in the design process: co-evolution of problem–solution. *Design Studies, 22*(5), 425-437.
French, R. M. (1995). *The subtlety of sameness: A theory and computer model of analogy-making*: MIT press.
Gabora, L., Rosch, E., & Aerts, D. (2008). Toward an ecological theory of concepts. *Ecological Psychology, 20*(1), 84-116. doi: 10.1080/10407410701766676
Gärdenfors, P. (2000). *Conceptual spaces: The geometry of thought*: Cambridge, MA, US: The MIT Press.
Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive science, 7*(2), 155-170.
Gentner, D., & Colhoun, J. (2010). Analogical processes in human thinking and learning *Towards a theory of thinking* (pp. 35-48): Springer.
Gero, J. S. (1994). Towards a model of exploration in computer-aided design. In J. S. Gero & E. Tyugu (Eds.), *Formal design methods for CAD* (pp. 315-336). Amsterdam: North-Holland.
Gero, J. S., & Fujii, H. (2000). A computational framework for concept formation in a situated design agent. *Knowledge-Based Systems, 13*(6), 361-368.
Gero, J. S., & Kannengiesser, U. (2004). The situated function–behaviour–structure framework. *Design Studies, 25*(4), 373-391.
Goel, A. K. (1997). Design, analogy, and creativity. *IEEE expert, 12*(3), 62-70.
Goel, A. K., Vattam, S., Wiltgen, B., & Helms, M. (2012). Cognitive, collaborative, conceptual and creative—four characteristics of the next generation of knowledge-based CAD systems: a study in biologically inspired design. *Computer-Aided Design, 44*(10), 879-900.
Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(3), 501-518. doi: 10.1037/0278-7393.11.3.501
Green, A. E., Kraemer, D. J., Fugelsang, J. A., Gray, J. R., & Dunbar, K. N. (2012). Neural correlates of creativity in analogical reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 264.
Greif, S. (1994). Computer systems as exploratory environments. In H. Keller, K. Schneider & B. Henderson (Eds.), *Curiosity and exploration* (pp. 287-306): Springer.
Gu, M. (2010). *Design of ART-based hierarchical clustering algorithm using quadratic junction neural networks.* Paper presented at the 2010 IEEE International Conference on Progress in Informatics and Computing (PIC).
Hawkins, J. (2005). *On Intelligence*. New York: Holt Paperbacks.
Hofstadter, D. R. (2008). *Fluid concepts and creative analogies: Computer models of the fundamental mechanisms of thought*: Basic books.
Holyoak, K. J. (1996). *Mental leaps: Analogy in creative thought*: MIT press.
James, W. (1890). The principles of psychology. *Harvard UP, Cambridge, MA.*
Jeong, C., & Kim, K. (2014). Creating patents on the new technology using analogy-based patent mining. *Expert Systems with Applications, 41*(8), 3605-3614. doi: http://dx.doi.org/10.1016/j.eswa.2013.11.045
Jupp, J. (2005). *Diagrammatic reasoning in design: computational and cognitive studies in similarity assessment.* (PhD), The University of Sydney, Sydney.
Kelly, N., & Gero, J. (2014). Interpretation in design: modelling how the situation changes during design activity. *Research in Engineering Design*, 1-16. doi: 10.1007/s00163-013-0168-y

Kelly, N., & Gero, J. S. (2011). Constructive interpretation with examples from floor plans. In C. M. Herr, N. Gu, S. Roudavski & M. A. Schnabel (Eds.), *Circuit bending, breaking and mending* (pp. 633-642). Hong Kong: CAADRIA.

Kennedy, P. J., & Shapiro, M. L. (2004). Retrieving memories via internal context requires the hippocampus. *The Journal of Neuroscience, 24*(31), 6979-6985. doi: 10.1523/jneurosci.1388-04.2004

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE, 78*(9), 1464-1480. doi: 10.1109/5.58325

Kruschke, J. K. (2011). Models of attentional learning. In E. M. Pothos & A. J. Will (Eds.), *Formal approaches in categorization* (pp. 120-152).

Langley, P. (1987). *Scientific discovery: Computational explorations of the creative processes*: MIT press.

Lenat, D. B. (1976). AM: An artificial intelligence approach to discovery in mathematics as heuristic search: DTIC Document.

Leopold, D. A., Wilke, M., Maier, A., & Logothetis, N. K. (2002). Stable perception of visually ambiguous patterns. *Nature neuroscience, 5*(6), 605-609.

Liapis, A., Martınez, H. P., Togelius, J., & Yannakakis, G. N. (2013). *Transforming Exploratory Creativity with DeLeNoX.* Paper presented at the Proceedings of the Fourth International Conference on Computational Creativity.

Long, G. M., & Toppino, T. C. (2004). Enduring interest in perceptual ambiguity: alternating views of reversible figures. *Psychological bulletin, 130*(5), 748.

Maher, M. L. (2000). A model of co-evolutionary design. *Engineering with computers, 16*(3-4), 195-208.

McCorduck, P. (1991). *Aaron's Code*: WH Freeman & Co.

McCrae, R. R. (1987). Creativity, divergent thinking, and openness to experience. *Journal of personality and social psychology, 52*(6), 1258.

Merrell, P., Schkufza, E., & Koltun, V. (2010). *Computer-generated residential building layouts.* Paper presented at the ACM Transactions on Graphics (TOG).

Mountcastle, V. B. (1997). The columnar organization of the neocortex. *Brain, 120*, 701-722.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 54.

Pachet, F. (2012). Musical virtuosity and creativity *Computers and Creativity* (pp. 115-146): Springer.

Phaller, K. A., & Squire, L. R. (2009). Biology of memory. In B. Sadock, V. Sadock & P. Ruiz (Eds.), *Kaplan and Sadock's comprehensive textbook of psychiatry 9th edition.* (pp. 665-666). Baltimore MD: Lippincott Williams and Wilkins.

Pylyshyn, Z. W. (1977). What the mind's eye tells the mind's brain: A critique of mental imagery *Images, Perception, and Knowledge* (pp. 1-36): Springer.

Qian, L., & Gero, J. S. (1996). Function-behavior-structure paths and their role in analogy-based design. *AI EDAM, 10*(4), 289-312.

Riecke, L., van Opstal, A., Goebel R, & Formisano, E. (2007). Sensory-perceptual transformations in primary auditory cortex. *The Journal of Neuroscience, 27*(46), 12684-12689.

Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., & Edwards, D. D. (1995). *Artificial intelligence: a modern approach* (Vol. 74). Englewood Cliffs, NJ: Prentice Hall

Schacter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 501-518.

Schneider, S., Fischer, J.-R., & König, R. (2011). Rethinking automated layout design: developing a creative evolutionary design method for the layout problems in

architecture and urban design *Design Computing and Cognition ’10* (pp. 367-386): Springer.
Schön, D. A. (1983). *The reflective practitioner: How professionals think in action* (Vol. 5126): Basic books.
Schon, D. A., & Wiggins, G. (1992). Kinds of seeing and their functions in designing. *Design Studies, 13*(2), 135-156.
Seelig, T. L. (2012). *inGenius: A crash course on creativity*: Hay House, Inc.
Shaffer, D., Hatfield, D., Svarovsky, G., Nash, P., Nulty, A., Bagley, E., . . . Mislevy, R. (2009). Epistemic network analysis: A prototype for 21st Century assessment of learning. *The International Journal of Learning and Media, 1*(2), 33-53.
Smith, B. D., & Garnett, G. E. (2012). Reinforcement learning and the creative, automated music improviser *Evolutionary and Biologically Inspired Music, Sound, Art and Design* (pp. 223-234): Springer.
Suwa, M., Gero, J., & Purcell, T. (2000). Unexpected discoveries and S-invention of design requirements: important vehicles for a design process. *Design Studies, 21*(6), 539-567. doi: 10.1016/s0142-694x(99)00034-4
Suwa, M., & Tversky, B. (1997). What do architects and students perceive in their design sketches? *Design Studies, 18*(4), 385-403.
Wiggins, G. A. (2006a). A preliminary framework for description, analysis and comparison of creative systems. *Knowledge-Based Systems, 19*(7), 449-458.
Wiggins, G. A. (2006b). Searching for computational creativity. *New Generation Computing, 24*(3), 209-222.