Interpretation in design: Modelling how the situation changes during design activity

Nick Kelly

Australian Digital Futures Institute, University of Southern Queensland, Australia

nick.kelly@usq.edu.au

John S Gero

Krasnow Institute for Advanced Study and Computational Social Science, George Mason University

Computer Science and Architecture, University of North Carolina at Charlotte

john@johngero.com

Abstract This paper presents computational implementations of models of the way that designers move between situations during design activity. The paper proposes a knowledge representation appropriate for situated models of design and describes the types of expectation present in such a model. Interpretation is described as a construction from expectations, and the movement from one situation to another as being influenced by these expectations. Implementations of this framework are described and three hypotheses about the way that designers move between situations are discussed. The implementations demonstrate the way that interpretation can lead to changes in the situation and the way that these changes are guided by past experiences. The paper discusses the significance of modelling the way that designers navigate their own knowledge and the role of changing expectations.

***Keywords**: situated design, computational modelling, interpretation*

Designers frequently change their notion of what they are doing during design, but how this happens has not yet been adequately explained, nor has it been modelled computationally. Design can be conceived as a sequence of actions, distinguishable activities over time that take place within a situation (J.S. Gero, 1998). A feature of designerly activity is that designers change their notion of what they are doing as the activity progresses (Cross, 1982; J.S. Gero, 1990). This paper presents a model, based upon the findings of protocol studies of the way that designers change their goals and design requirements during design activity (Schon & Wiggins, 1992; Suwa, Gero, & Purcell, 2000), that can support computation of this type of designerly activity. Design is taken to be a situated activity and knowledge as grounded experience. From this perspective a description of movement between situations is articulated and a computational model developed, implemented and used to examine hypotheses about designers moving between situations.

The term *situation* (Clancey, 1997) refers here to a designer's internal notion of the world at a particular time. This definition of situation in design is related to terms that emphasise aspects of the same notion: the current *internal context* (Kennedy & Shapiro, 2004), the current *epistemic frame* (Shaffer et al., 2009) and the current *ecology of mind* (Gabora, Rosch, & Aerts, 2008)[1]. A designer holds grounded knowledge from experience within the world (L W Barsalou, 2007). The situation refers to the parts of this knowledge that are being used to understand the world at this current moment and in this current experience; useful terms to describe these parts of knowledge are concepts and percepts. At a particular moment in time the designer holds a view of the world that comes from experience – from those parts of experience relevant to this moment. The interpretation that a designer makes is shaped by the current situation, i.e., this particular formulation of knowledge from experience. The work in this paper describes this interaction between experience and interpretation and shows how it leads to hypotheses about the way that designers move between different situations during design.

The motivations for conceptualizing design as a sequence of situated acts are both to explain aspects of design that have not been adequately explained and to suggest ways that computation can become more designerly. There is empirical evidence that designers' notions of what they are doing change during design activity (Fish & Scrivener, 1990; Gross & Do, 2000; Suwa & Tversky, 1997). There is not yet an adequate understanding of the cognitive processes that relate knowledge from experience to the changing conception of the design task, nor of how to satisfactorily model them. An example of this is the way that designers make unexpected discoveries within their own work that inform their future design actions (Suwa, et al., 2000). These designerly phenomena are addressed in this paper through a situated framework (Clancey, 1997; J.S. Gero & Fujii, 2000). The paper develops a relationship between first person knowledge and interpretation within a situation. It addresses the questions of: (i) what does it mean for a designer to be in a situation; and (ii) how do designers move from one situation to another.

[1] Gero and Smith (2009) focus upon these distinctions of definition.

These questions are defined then explored through the notion of situated interpretation. The paper describes an approach to knowledge representation suitable for modelling situated design. This knowledge representation is used first to describe the role of expectations in a situation and then to present a model of interpretation in design showing the interaction between expectations and the external world at different levels of abstraction. This model leads to specific hypotheses about design activity. The work uses computational models of interpretation to examine the hypotheses and to discuss their implications for design.

1 Background

1.1 Interpretation and unexpected discovery in design

Whenever a designer brings something from the external world into their internal world, interpretation is occurring. For example, when a designer reads a design brief, considers their own work whilst sketching, or observes behaviours of a model, interpretation occurs – something external is given meaning internally by the designer. The literature reflects this movement from interpretation of a source AS a concept (Z. W. Pylyshyn, 1977) towards interpretation as situated conceptualisation (Lawrence W Barsalou, 2009).

The interpretation that a designer produces arises from the interaction between the source (what it is that the designer is interpreting), the designer's previous experiences (the knowledge held by the designer) and the situation (the world-view made up from parts of this knowledge). This can result in a designer being able to interpret the same work in many different ways when in different situations; this is captured in Henry David Thoreau's (1851/1993) statement that "it's not what you look at that matters but what you see".

A designer draws upon knowledge from past experiences. This knowledge is made use of within the current way of viewing the world (the situation); and as design progresses, this situation changes (John S Gero & Kannengiesser, 2004). This way of viewing design is as a progression of changing situations within which actions occur. The motivation for adopting this perspective is to understand the cognitive processes of changing interpretations of the world that appear to occur during design.

1.1.1 Unexpected discovery

A phenomenon in design that illustrates this change of situation is that of *unexpected discovery*, the invention of design issues and requirements during design activity (Suwa, et al., 2000). When designers look at their own work they are capable of inventing new design issues and requirements, as observed by Gombrich (1966) that "in searching for a new solution Leonardo [da Vinci] projected new meanings into the forms he saw in his old discarded sketches". Studies related to this phenomenon (Bilda, Gero, & Purcell, 2006; Menezes & Lawson, 2006; Schon & Wiggins, 1992; Suwa, et al., 2000) have typified the kinds of discoveries that are made by designers and linked some of these to the situation of the designer. Suwa et al (2000) note that whilst some functions arise based upon the list of initial requirements, others seem to be invented during design activity, such as those that are directed by the use of explicit knowledge or past cases, or those that have been extended from a previous goal through concretizing or broadening. Designers produce new design issues and requirements that are specific to their own experiences, that point in time, the state of the design medium, the state of the collaborators and the state of the beliefs about the design task. The new design issues and requirements indicate a change of situation. In this work a model is articulated for the kinds of cognitive processes that might explain how a designer moves from one situation to another, such that to an outside observer the designer appears to have invented a goal or requirement for the design.

1.2 First person knowledge

The central notion of situated design is that knowledge can be grounded in experience leading to knowledge that can be described as *first person knowledge*. There is no universally agreed definition of knowledge. A commonly accepted useful definition of knowledge is that it is a well-grounded belief with evidentiary support that is independent of the proponent of the knowledge. This independence of knowledge from its proponent lays the foundation for objective knowledge in science. For example, Newton's laws, once they have been propounded, do not depend on the existence of Newton. This kind of knowledge is called "third person" knowledge in that a person removed from the original proponent of the knowledge can understand and make use of it. This is in contradistinction to knowledge that individuals gain through their experience and that is not independent of them and their experiences and is called "first person" knowledge. Much of the lives of individuals are built around first person knowledge rather than third person knowledge.

A movement towards knowledge that is situated is a movement towards first person knowledge. In the use of knowledge in design this is captured by Gero's (2007) principles of a computational paradigm for situated design:

1. **Principle of Effect:** What you do matters. Actions that the system takes can change the knowledge held by the system.
2. **Principle of Ordered Temporality:** When you do what you do matters. The chronology of the system's use affects what is used and what is learned.
3. **Lemma of Experience:** What you did before affects what you do now. A consequence of principles 1 and 2.
4. **Principle of Locality:** Where you are when you do what you do matters. The state of the world external to the system affects the use of knowledge.
5. **Principle of Interaction**: Who and what you interact with matters. A system learns through interaction with users and other systems.
6. **Principle of Ontology:** What you think the world is about affects what it is about for you. A system's conception of the world changes the interpretation of the world

This is a call for systems that hold first person knowledge, where the learning and re-use of knowledge is tied to the experiences of the design agent:

> "It is claimed that situated design computing contains the seeds for the development of computational constructs that can be used to produce systems that match our notions of designerly activity. The two ideas of situatedness and constructive memory provide the basis for the development and use of experience in alternate situations. This is not a form of case-based reasoning or of default reasoning. It is a novel approach based on foundational concepts from cognition." Gero (2007).

When an external observer sees a designer at work, there is a changing conception of the goals of the design activity – they are a part of the situation of the designer. As design activity progresses the situation is seen to change as the designer brings new design elements to the design. An example of this is observed in a cognitive study of designing using a think-aloud protocol:

> "Our architect once decided to bring a water stream from the open plaza in front of the museum building into the entrance hall, as a means to guide visitors into the building in a cheerful way. Then, after a while, he noticed that water in the building may cause problems because humidity affects the artworks. But, because he thought that the idea of bringing water inside is still promising to produce a lively atmosphere, he set up a goal to search for a method to let artworks and water co-exist." (Suwa et al 2000)

The designer has commenced this activity within one situation, wanting to bring water into an open plaza as a guide. In interpreting the consequences of this when expressed in the design medium, the situation changes, bringing in a new goal to find a way of having the water and the artworks co-exist. This is distinct to a designer identifying a subgoal of an existing goal (Newell & Simon, 1961). In drawing from experience during interpretation (observing the possible problems of humidity) the designer has invented a new goal (of having artworks and water together without the problems of humidity) that is a part of a changed situation. The work described in this paper comes from a motivation to understand how these changes to situation come about and how it is possible to model them.

1.3 Outline of the paper

Design is taken to be a sequence of situated acts and the notion of unexpected discovery has been used to give an example of the link between experience and interpretation. The remainder of the paper is concerned with: (i) describing an approach to knowledge representation suitable for a situated design agent; (ii) describing the expectations held by such an agent; (iii) presenting a model for interpretation in such an agent; and (iv) hypothesising the emergent behaviours that can be expected to be observed in such an agent during design activity. After explaining the basis for the hypotheses the paper describes three implementations that demonstrate how the effects of these hypotheses can emerge from experience, expectation and interpretation.

2 How designers move between situations during design activity

2.1 Representing knowledge for design

The following question was posed by Schön and Wiggins (1992): How do designers change the way that they see their own work during the design process? This research begins by positing the cognitive processes involved in interpretation and by exploring the connection between the first-person knowledge held by a designer and the types of seeing observed during design activity. It asks how designers make use of knowledge from experience when looking at their own work.

Any attempt to model interpretation computationally requires a computational model of knowledge. The focus of this work is upon representing knowledge for design that includes first-person knowledge that is held by a designer and utilised within a situation. In representing knowledge in a way that is suitable for modelling situated design the same experiences can be utilised differently in different situations (Peng & Gero, 2006). Two notions that are relevant for such a system are *abstraction*, the creation of summary representations (L W Barsalou, 2005) and *expectation*, the notion that what an agent expects within a situation changes what they see. A framework is outlined here for a concept formation system for design that is driven by expectation and develops knowledge at different layers of abstraction. The work has its basis in the notion of *perceptual symbol systems* (Barsalou 1999).

In perceptual symbol systems cognitive representations are based upon perceptual states (L W Barsalou, 1999). This is in contrast to an amodal symbol system, in which cognitive representations use a language which does not have a perceptual basis, e.g. natural language or abstract symbols (Fodor, 1975; Minsky, 1974; Pylyshyn, 1984). Barsalou (1999), Gärdenfors (2000), and Hawkins (2005) all propose different frameworks for how a perceptual symbol system can be modelled, and these types of models have been successfully used to computationally model specific phenomena (Pezzulo et al., 2011). In these models, concepts arise as summary representations over perceptual information. What starts as information in each of the senses becomes abstracted to the level where it can be manipulated as symbols, a transition from sub-symbolic to symbolic (Harnad, 1999). The cognitive capabilities of a system arise from a combination of its native functions as well as its experiences in a reconciliation of nativism and constructivism (Karmiloff-Smith, 1994).

In this work the *conceptual spaces* paradigm (Gärdenfors, 2000) is used to construct a computational model because: (i) it has been demonstrated to be useful for computation (Beyer, Cimiano, & Griffiths, 2012; Gärdenfors, 2000); and (ii) with the addition of situations to the model it becomes a computational framework that is a fit with the principles for situated design computing. This framework is based on the one outlined by Gero and Fujii (2000).

There are four levels of abstraction in representing the knowledge that makes up this framework. The reasons for identifying four levels are as follows. At the lowest level of abstraction, a way of describing the data from the sensors of an agent is required, referred to here as *senscepts* (Montare, 1994) for uniformity. At the highest level of abstraction, agents learn how to co-ordinate their knowledge from experience (Clancey, 1999) and these highest level constructs will be referred to as *situations*. Between these sensor notions and the highest level of abstraction, an agent holds knowledge at different levels of abstraction. Rather than trying to accurately model humans, this work is concerned with creating a representation that is sufficient for modelling aspects of design. The structure is simplified by following Gärdenfors (2000) and using the notions of *percepts* and *concepts* to represent these layers of abstraction between senscepts and situations rather than weighing into the question of exactly how this fluid hierarchy of knowledge is structured. These four levels of abstraction of knowledge make use of the conceptual spaces architecture following Gärdenfors and an extension to include situations.

2.1.1 Conceptual spaces: senscepts, percepts and concepts

An agent has sensors which produce data during interaction with the external world. For example, an eye can sense changes in light, an ear changes in air pressure. Each datum from a sensor interacting with the world is referred to as a *senscept*. A senscept implies a *dimension*. For example, a human eye might generate a senscept for each of hue, brightness and saturation at a point in time (Gärdenfors, 2000). A dimension is the one-dimensional space within which the senscepts are located. Each dimension is a continuous variable of a certain range.

Dimensions that are inseparable create a *perceptual domain* within which *percepts* are located. Inseparable means that the agent cannot get information for one of the dimensions without getting information for all of them. To continue the previous example: in the human eye, hue cannot be obtained without also obtaining brightness and saturation (Gärdenfors, 2000). A perceptual domain (e.g. texture, colour) is a space with dimensions of those things that the agent can sense. In the example of colour being sensed by hue, brightness and chromaticity, this results in a three-dimensional domain. Senscepts from a stimulus are grouped together to create a percept represented by a value within this perceptual domain. A percept within a perceptual domain is a vector space arising from the perceptual dimensions that constitute it. Just one layer is described here as indicative of all layers.

A *concept* is a convergence zone that brings together spaces within perceptual domains that experience has shown to be related. For example, the concept for *banana* might bring together areas in the colour domain that might be called yellow, green and brown with areas in a shape domain that are associated with the Lady Finger and Cavendish banana varieties. It implicitly adopts a prototype theory of concepts where the most typical perceptual regions (e.g. the colours and shapes above for a *banana*) are associated and less typical instances of a concept have some distance (measured in conceptual space) from a prototype (Murphy, 2002). Figure 1 shows the way that dimensions create the space of a perceptual domain, and the way that a concept associates regions in domains with each other. A part of the meaning of a concept comes from its relationship with other concepts. Just one layer is described here, although humans hold concepts at many layers of abstraction (Rosch, 1978).

The strength of this formulation is that there is now a geometric representation of senscepts, percepts and concepts. Within this space, similarity can be calculated as the distance between two points within conceptual space (Gärdenfors, 2000; Nosofsky, 1988). This is important for interpretation, where similarity to existing concepts plays a role in determining whether the expectations are able to account for what the agent is experiencing.

2.1.2 Adding situations to the knowledge representation

Situations can be thought of as a convergence zone for these concepts – the memory of the co-ordination of concepts (L W Barsalou, 1999; Clancey, 1997, 1999). Concepts that are utilised at the same time within an agent are a part of the same situation. This framework is extended by introducing a situation as a collection of concepts. The utility of a situation is that it allows for the explicit representation of the world view that arises from the use of a particular network of concepts.

Situation is different from the other layers because it alone does not receive expectations from a layer above it. For example, perception knows what to expect of sensation because conception passes expectations down to it. In contrast, situation relies upon grounded experience alone.

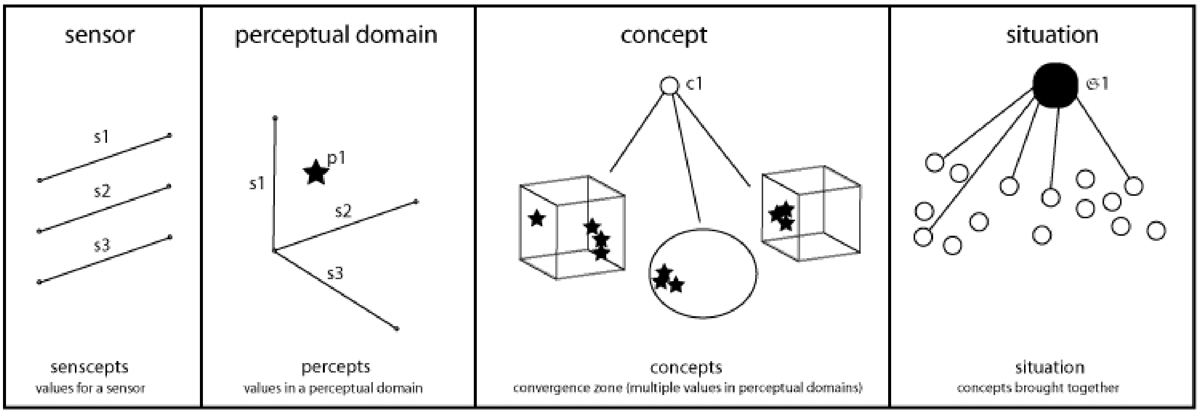

Figure 1 Levels in a hierarchy of knowledge organisation for design

2.2 Implicit and explicit expectations of a designer

This language of knowledge representation is used to describe the expectations of a designer during design activity. A designer holds knowledge about the world from their experience. At a certain moment in time, some of this knowledge is explicitly being utilised: asking the question “What are you doing now?” (Clancey, 1999) or think-aloud studies (Wright & Monk, 1985) are one way to get a sense of the explicit expectations of a designer.

One attempt at describing the explicit expectations of a designer engaged in design activity is in the notion of *design prototypes* (J.S. Gero, 1990), which suggest that a designer holds notions about function, structure, expected and actual behaviour of the design; relational knowledge; qualitative knowledge; computational knowledge; and context knowledge. Moving from the construct of a design prototype back to the notion of a perceptual symbol system, these notions of the design can be described as types of concepts, percepts and senscepts within a situation that are being used as resources by the designer in the current activity.

In addition to this explicit knowledge within the situation there is *implicit* knowledge from the situation within the internal context of other knowledge from experience. An agent creates connections between different layers within the knowledge representation through experience as part of what is termed ‘memory’. Only some of these connections are utilised as a part of the explicit expectations. Implicit expectations, what might be described as connections that are below a liminal threshold, are those connections which exist but of which the agent is not aware. In these models of movement between situations these implicit expectations play a role in how the designer moves from one situation to another, and may be a guide to starting to answer the elusive question of: Why this design move and not another?

The suggestion is then that a designer holds what can be described as two different types of expectations, those of which they are aware, the explicit, and those of which they are not aware or that are only implied by the structure of knowledge, the implicit. There are three notions from the literature that relate to attempts to model the effects of expectations: explicit and implicit memory – which map directly to explicit and implicit expectations; priming – how implicit expectations affect human behaviour; and selective attention – how expectations change human experience of stimuli. We provide a very brief introduction to these three notions for those not familiar with them.

2.2.1 Explicit and implicit memory

Explicit memory is the form of memory used “when performance on a task requires conscious recollection of previous experiences”(Graf & Schacter, 1985). *Implicit memory* is “when performance on a task is facilitated in the absence of conscious recollection" (Graf & Schacter, 1985) or “a behavioural, emotional and perceptual form of memory devoid of the subjective internal experience of self or of past" (Cozolino & Siegel, 2009). An example of explicit memory is deliberately bringing to mind the appearance of your first pet. An example of implicit memory is thinking that a statement is true because you think you have overheard somebody making that statement in the past; where, crucially, you are unaware of the use of your memory of hearing that statement (Begg, Anas, & Farinacci, 1992).

2.2.2 Priming

Priming is an effect observed when exposure to a stimulus affects the response to a subsequent stimulus. Experiments have demonstrated both perceptual priming (of forms) and conceptual priming (of categories) (Schacter, 1987). For example, “words might be presented in a study phase and then again, after a delay in a test phase when a priming measure such as reading speed is obtained. Participants are instructed to read words as quickly as possible in such a test and they are not informed that memory is being assessed. It is observed that priming improves the speed of recognition" (Phaller & Squire, 2009).

2.2.3 Selective attention

Selective attention is where certain features of either external stimuli or internal representations are noticed at the expense of others. Selectivity of attention has three dimensions (Cozolino & Siegel, 2009): (i) filtering - focussing on specific attributes (e.g. large shapes rather than small shapes); (ii) categorising - recognising information based on stimulus class (such as attending to squares and not other shapes); and (iii) pigeonholing - only using perceptual information needed to place a stimulus into a category (e.g. only looking at number of sides of a shape to classify it).

2.3 Interpretation within a situation

From these notions of a knowledge representation in which expectations are both implicit and explicit, a notion of what interpretation looks like in a situated design agent can be developed. Interpretation is concerned with reconciling two opposing goals of an agent: (i) to maintain a stable notion of the world; and (ii) to change its picture of the world when the world changes. Useful words to describe these opposing goals are “assimilation” (changing data from the world to fit expectations) and “accommodation” (changing expectations to fit the external world) after Piaget (1954). In describing a cognitive architecture that fits with design these two goals are described by the direction in which data flows. The word *pull* describes the flow of expectations – an attempt to construct what is expected from what is available. The word *push* describes the flow of data from the external world into perception and up the hierarchy of abstraction, challenging expectations.

Interpretation is defined here from a situated cognition perspective as a continuous, dynamic, constructive activity that attempts to construct an internal representation from a source, using expectations where it is possible and constructing an explanation from existing knowledge where it is not. It is an activity that is distributed across layers, with each layer attempting to first pull from the data available using expectations, and subsequently push from the data where this is unsuccessful.

In design, this push and pull can be observed in the conversation with a design medium such as a sketch (John S Gero & Kannengiesser, 2004). The external source is interpreted using expectations – for example, the ideas held by the designer when making the sketch. Changes in the situation can be described as occurring when something does not adequately fit the expectations used for interpretation. Cases can be described where something that the designer is interpreting does not fit the expectations at the different levels of sensation, perception, conception and situation and how something might be constructed that does fit.

Interpretation begins with pull, an attempt to construct what is expected from what is available. The output of pull is interpreted data of the type described by the layer. The interpreted data from pull in sensation, perception, conception and situation respectively are senscepts, percepts, concepts and situations. Pull tries to construct expectations from the source for the layer. Push is the part of interpretation that begins with the source. There will be times when expectations are not able to be constructed, when there are incorrect expectations or when something new is encountered. Push is the part of interpretation that allows for the unexpected to be recognised and allows for new knowledge to be learned.

An example can be used to show the difference between push and pull. A stimulus can demand attention, `pushing' its way into sensation, such as the way that regardless of your expectations when your mobile phone rings, you notice it. Pull shows the way that expectations can change an experience. Consider that you are attempting to meet up with a friend in a crowd. Based upon the back of somebody's head you think that you see your friend and rush to catch up with them. However, as you get closer you realise that they look only a little like your friend. This is an example of pull because the expectations of seeing the friend have changed the way that the world is seen: the interpretation has been constructed from expectations.

The term source is used because it recognises that from any source it is possible to pull many different kinds of interpreted data. It is possible to internally represent a single source in many different ways. For example, a stimulus will yield different senscepts depending upon which sensors are pointed at it and how those sensors are primed. An ear pointed at a lawnmower and an eye pointed at the lawnmower produce different senscepts, and an ear expecting the lawnmower produces different senscepts to one that is not. The same construction from the source is present in all of the layers. A layer in a system of interpretation is shown in Figure 2.

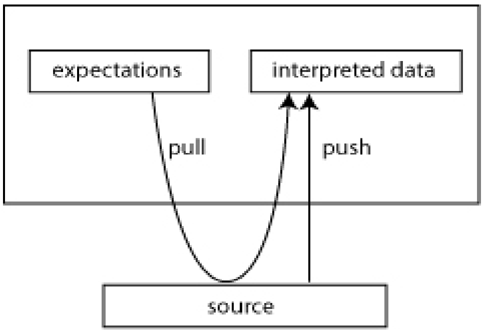

Figure 2 Pull from expectations and push from data within a layer

2.3.1 Examples from each layer during interpretation

Sensation pulls data from the source based upon what it expects to sense, through control of its sensors. Expectations of sensation during interpretation are driven by the requirements of perception. For example, people turn so that their eyes are facing towards something that they want to see, or act to touch something that they want to feel. As well as controlling the orientation of their sensors people control the functioning of their sensors, such as the opening of an eyelid or the flaring of nostrils.

Perception pulls from senscepts to construct percepts, based upon what it expects to see. Expectations of perception during interpretation are driven by the requirements of conception. Gestalt perception provides examples of perceptual constructions based upon expectation (Zusne, 1970). An agent constructs the illusory Kanisza triangle seen in Figure 3 by the continuation of the lines suggested by the sectors cut out from the three black circles. In another example, people are played an audio track, consisting of a continuous musical note interrupted by white noise before the note resumes. People are able to hear the note continuing through the middle of the white noise (Riecke, van Opstal, Goebel R, &

Formisano, 2007). This is an example of perception in the auditory system pulling out of the noise what it expects to be finding.

Conception pulls from percepts to construct concepts, based upon what it expects. Expectations of conception during interpretation are driven by the requirements of situation. There are two aspects to pull from conception: (i) an attempt to pull percepts that are expected to occur within the expected concepts; and (ii) changing the use of percepts to fit the concepts. Examples of pull from conception can be seen in conceptual priming experiments (Schacter, 1987), such as the one performed by Graf et al (1985) where subjects were primed with a list of words and performed a task of identifying eight exemplars for a category, where it was demonstrated that the primed words had influenced the spoken words.

Situation pulls from the available concepts to construct the expected situation. Unlike the other layers, situation does not have a layer above it. When the expected situation cannot be constructed from the available concepts the situation changes either through: (i) a switch to another situation that experience suggests will be useful; or (ii) a change to the current situation. Through experience, agents come to expect concepts to be co-ordinated in a certain way. There are two aspects to pull from situation: (i) an attempt to pull concepts that are expected to occur together; and (ii) changing the use of concepts to fit the situation. An example of this can be seen in the `runway width illusion' encountered by pilots, Figure 4 (Parmet & Gillingham, 2002). In this illusion, the pilot is accustomed to a certain runway width. When approaching a runway that is narrower than they are accustomed to, the pilot thinks that they are higher than they really are, scenario (b) in Figure 4. In terms of pull, the pilot is expecting the runway on approach to have certain visual characteristics. The perceptual expectations of the pilot are being met, but this is only because of the frame of expectations within which they are operating. If an instructor sitting next to the pilot were to tell them that the runway was narrower than they expected, then the concepts available to situation no longer fit with the concepts expected. As a result, the situation changes to another situation. Within this new situation, exactly the same perceptual data leads the pilot to a safe landing.

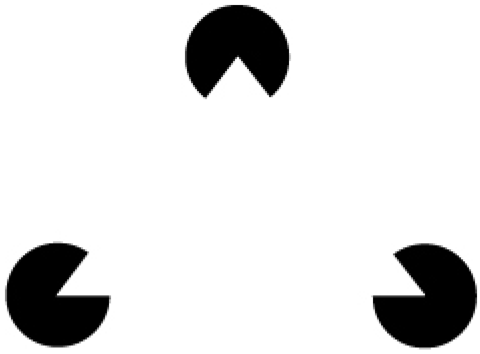

Figure 3 A Kanisza triangle

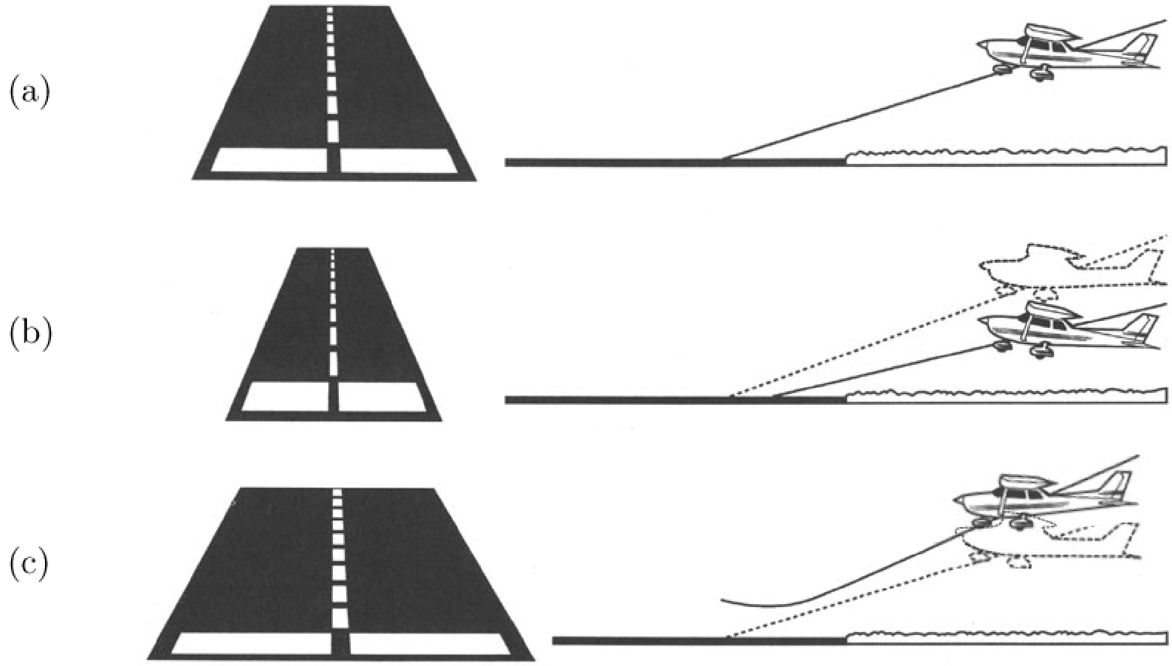

Figure 4 The effect of runway width on the pilot's image of the runway (left) and the potential effect on the approach path flown (right) where dashed line indicates pilot perception: (a) normal approach; (b) narrow runway makes the pilot feel higher than they actually are; and (c) wide runway makes the pilot feel lower than they actually are (after Parmet and Gillingham 2002)

2.3.2 Interpretation as dynamic and constructive

Interpretation in this way forms a part of a complex system in which small departures from expectations can lead to large changes in the world view of the agent. Figure 5 shows a dynamic system constructed from multiple self-similar layers of the type seen in Figure 2. Each layer attempts to construct what it expects to find from what is available in the layer below. The output from each layer is to make data available to the layer above, and to feed back to the layer below to alter its expectations. It is difficult to predict the changes that will occur in the system when one layer is changed, where a small change in the topmost layer can lead to large changes in the layers below, or that a small change in a lower layer can result in all layers changing. This is inspired by the models of Hawkins, Barsalou and Gärdenfors, who are in turn inspired by the neuroscience literature (L W Barsalou, 1999; Gärdenfors, 2000; Hawkins, 2005; Mountcastle, 1997).

This phenomenon has been observed in designers changing their design trajectory through observations within their own sketches (Suwa, et al., 2000; Suwa & Tversky, 1997) where small changes in a representation can result in a re-conceptualization of the design problem.

During interpretation, if a layer cannot construct what it is expecting to find from what is available, it changes its expectations. This idea of designers moving from one concept to another or from one situation to another situation is the focus of this work, i.e., *What guides this movement from one expectation to another expectation?*

The response introduced and explored in this work is to propose that *connections within and between layers at different levels of abstraction are developed through experience; and that the movement to a different expectation is guided by these connections and the data available.* Through interaction with the world over the course of its life, the agent has developed connections within a layer (e.g. between concepts and concepts) and between layers (e.g. between concepts and percepts). The hypothesis is that the design conversation with a medium is a way of navigating this type of knowledge.

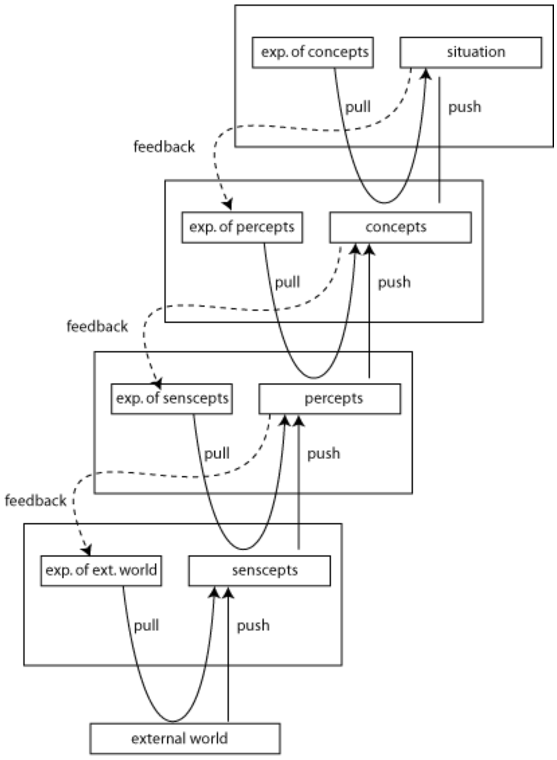

Figure 5 The dynamic system resulting from multiple layers with the presence of feedback from the interpreted data of one layer to the expectations of the layer below

2.4 Movement between situations.

An approach to knowledge representation has been described that is both a good fit for situated design and applicable for computation, using the language of senscepts, percepts, concepts and situations. Within this model designers can be described as holding both explicit and implicit expectations. Interpretation is described as an activity that attempts to construct from expectations at different layers of abstraction. Expectations are changed using grounded knowledge from experience when existing expectations are no longer useful.

Interpretation of a source occurs within a situation and produces an internal representation. From this description of interpretation as a dynamic, constructive activity of push and pull from what is available, the following three hypotheses about designers interpreting can be proposed that arise:

1. **Attempts to construct an interpretation using expectations can lead to a change in the situation.** When pull is unable to produce what is expected, expectations can be changed by push during interpretation; this is the initial phenomenon described by Suwa et al (Suwa, et al., 2000; 1997). The hypothesis is

that this movement to a changed situation can occur as a part of the activity of interpretation, as described in this paper.

2. **Small changes in a source can lead to large changes in the internal representation.** If the proposed dynamic model of interpretation at different levels of abstraction fits then we expect to observe this type of phenomena in designers where small changes to a source, through the complex system that draws upon the experiences of a designer, can manifest as large changes to the internal representation.

3. **Changes to the situation have their origin in the experiences of the design agent, manifested as connections between knowledge at different levels within a hierarchy of abstraction.** In answering the question "where do design expectations come from?" there is the possibility of making a connection between experience (how it is held by a designer) and the cognitive processes by which experience brings expectations into a design activity.

3 Modelling movement between situations during design activity

Three computational models were produced based upon a conceptual spaces knowledge representation, in which aspects of the system are situated and interpretation occurs as a dynamic construction through push and pull.

3.1 The approach

Self-Organising Maps (SOMs) (Kohonen, 1990) facilitate unsupervised learning as well as having a straightforward learning algorithm, making them suitable for adaptation to a hierarchy of SOMs (Dittenbach, Merkl, & Rauber, 2000) as called for by the model of knowledge representation for a situated agent, as well as suitable for experimenting with feedback and feedforward within the hierarchy.

The models described in this section use two layers of linked SOMs in the testing of the hypotheses. Figure 6 shows two connected SOMs, where the activation of nodes in the lower layer forms the input into the higher layer, marking an abstraction from the data. The lower layer creates an abstraction from input data whilst the higher layer creates an abstraction over the activity of the lower layer.

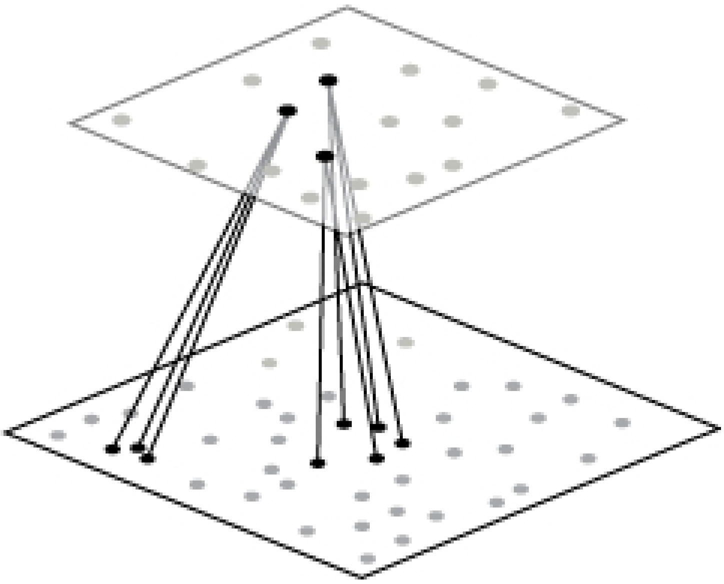

Figure 6 Linked SOMs representing abstraction from data.

Interpretation within the linked network uses both pull (from expectations, with feedback through the network) and push (from data, with feedforward through the network).

An overview of push and pull within a two layer neural network as a sequence of steps is presented in Figure 7. Step 1 in Figure 7 is an attempt to pull from expectations in layer 1. If pull cannot occur, then step 2 is push from the source. Either from push or pull, the interpreted data from layer 1 is the data available to the layer above, layer 2. Step 3 is an analog for step 1, where layer 2 attempts to pull from expectations using the data available from layer 1. If the expectations of layer 2 cannot be used for pull, then push into layer 2 occurs, step 4. This is different to push in layer 1 because it is the topmost layer and there are no higher layers in which to test expectations. Push in layer 2, step 4, results in either: (i) a change to different expectations that are more suited; or (ii) the construction of a new expectation based upon the data.

If the expectations change in layer 2 then the expectations of layer 1 are updated as feedback, step 5. Step 1 is now repeated, but with new expectations, step 6. This cycle can be iterated and extended to multiple linked networks.

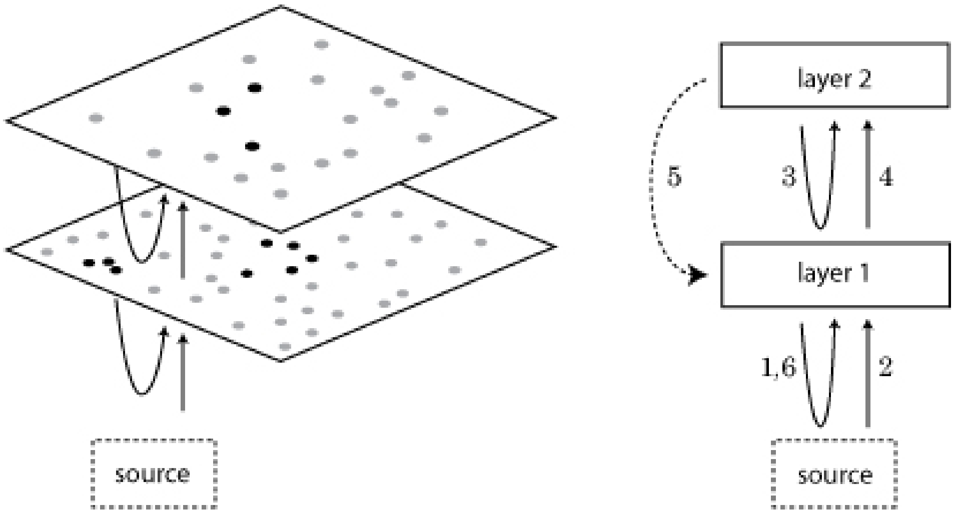

Figure 7 Push-pull in a two layer system with numbered steps of: (1) pull from expectations in layer 1; (2) push from the source into layer 1; (3) pull from the expectations of layer 2; (4) push from layer 1 into layer 2; (5) an update of layer 1 expectations by layer 2; (6) interpretation through pull from layer 1 with the updated expectations.

3.1.1 Describing the implementation

Two SOMs are implemented to represent a lower layer of abstraction SOM^1 and a higher layer of abstraction SOM^2. Each is a 2D SOM described by the number of nodes in each dimension, u and v, and a number of features, f, Equations 1 and 2.

$$SOM^1(u, v, f) \qquad (1)$$

$$SOM^2(u, v, f) \qquad (2)$$

During training, each SOM uses the Kohonen training algorithm (Kohonen, 1990) in a two phase training of: (i) reducing the neighbourhood radius to 1; and (ii) reducing the learning increment from 0.1 to 0. If an input vector of x is introduced to the network then the best matching node $BMU = m_{u,v}$ is the one that has the least Euclidian distance $d_E(x, m_{u,v})$.

In training, the outputs from SOM^1 form the inputs into SOM^2 with co-ordination of multiple outputs forming a single input in some of the models described. The situated effects described here only take effect following training.

In a typical SOM the set of all nodes $M = \{m_{0,0}, m_{0,1}, m_{02}, \dots, m_{1,0}, \dots, m_{u,v}\}$ is used to establish BMU. In the models described here expectations are implemented as a subset of all nodes.

$$\{M^* | M^* \subset M\} \qquad (3)$$

The similarity in the models is defined by the Nosofsky distance (Nosofsky, 1988) between the input and best matching unit $d_N\left(x, m_{u,v}\right) = e^{-c*(d_E)^2}$. A similarity threshold σ is defined as a minimum similarity required for pull to occur in the system. Pull can only occur within a SOM if the inequality described in Equation 4 holds.

$$\{BMU \in M^* | d_N(BMU, x) < \sigma\} \qquad (4)$$

When the set described by Equation 4 is empty in SOM^1, a cue is given to SOM^2 that a change of expectations is required. This comes from the layer above. If this is the top layer then the BMU here is used. The change to $M*$ is in this way a function of the output from the layer above.

3.2 Situations and expectations

In section 2.2 two types of expectation were described as explicit and implicit; those that are a part of the situation and those that are implied by the knowledge structure surrounding the situation. A further categorisation of expectations into those that are *spatial* and those that are *temporal* is useful for describing the modelling of expectations. Spatial expectations relate to the idea that observed co-occurring instances will occur together again. Temporal expectations relate to the idea that an observed sequential progression will occur again. This section describes a demonstration using shapes, and then presents two variations that were carried out with an environment changing over time and then a similar experiment with a different environment of letters.

3.2.1 Changing situations in a constructive interpretation agent

Two linked self-organising maps as seen in Figure 7 were trained on a set of representations generated pseudo-randomly by an algorithm. Each representation was made up of three shapes in a linear sequence, with each shape overlaid upon a tartan grid. The network was initially trained on 500,000 such representations. The rules for generating each representation in the training phase were:

(i) Each of the three tartan grids is 16 x 16 squares in size

(ii) The first square is filled with a randomly determined shape. This means randomly selecting one of three algorithms for filling in black squares in the tartan grid to create one of three possible shapes: a square, a cross or a triangle. Each algorithm allowed for a random width and height

(iii) The other two squares are subsequently filled out to produce one of two patterns: (a) with all three squares the same shape, although potentially different sizes through random generation; or (b) with all three shapes different, again with potentially different sizes.

The representation at *t*=1 in Figure 8 gives an example of a representation of three squares produced in this way. The rules described above in training are manifested in the network as what can be described as expectations that could be inferred by a human observer as a heuristic, e.g. if it sees two shapes the same it will have expectations that the third shape will be similar.

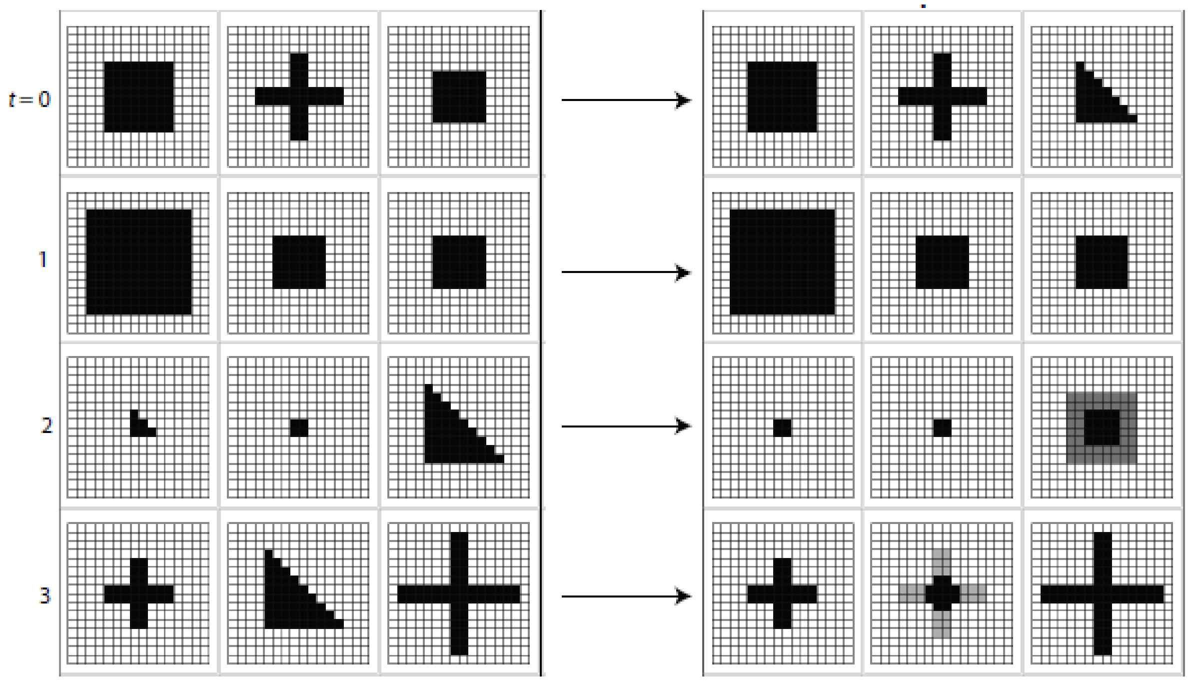

Figure 8 Four representations (left block) and internal representation after interpreting (right block) as a time sequence (the light grey is an artefact of the abstraction that occurs during training of the SOM).

The first model demonstrates the capacity for a computational model to learn concepts from experience in a way that is situated, and to then demonstrate how this conceptual co-ordination changes the way the world looks when constructively interpreting. The notion of a threshold of similarity is used to designate the degree of similarity that is required to be able to construct from an expectation using the data available. This degree of similarity required is dynamic and changing with the situation.

In the test phase of the model following training, three random shapes were produced to create a representation that may or may not obey the rules presented during training, e.g. it could be two squares and a triangle. The way that constructive interpretation occurs in the model can be described by:

(i) **Temporal explicit expectations:** The trained network holds expectations that the shapes seen at time *t* will still be present at time *t+1*. For example, if it has seen three squares at time *t* it will attempt to construct three squares at time *t+1*. If the similarity threshold is not satisfied at each level of the network (firstly in constructing each of the three shapes found as squares in layer 1, then in putting the three shapes presented to layer 2 together) then the expectation needs to be changed

(ii) **Spatial implicit expectations:** When interpretation in a layer is unable to construct what it is expecting to find, the network changes its expectations. The knowledge structure present in the agent from training guides this change of expectations. In this case, this is a spatial expectation that if, for example, one square is found then it will expect to find two more squares or two different shapes. This is a movement that is guided by the proximity of nodes within the network of SOMs.

(iii) **Learning:** In interpretation there will be occasions where neither explicit expectations in the situation, nor implicit expectations from a guided search of knowledge, can adequately construct an explanation for the data that a layer is finding. In this case, learning would occur. However, this model did not implement a learning algorithm in the testing phase, rather it progressively reduces the similarity threshold until an interpretation results.

An example of the results produced by the constructive interpretation model is shown in Figure 8. Two notable effects of constructive interpretation occur at times *t*=2 and *t*=3 where the model produces an internal representation different to what a human observer might see. At time *t*=2 the model has an expectation of seeing three squares. This expectation is strong enough, or the data is similar enough, that the lower layer can construct what it is expecting from what it sees and it follows that the higher layer can do the same. In this case the temporal expectation has been used to construct the internal representation. The case at time *t=3* is different, as the temporal expectations have not been useful – there is not sufficient similarity to construct from these expectations. In this case, the network uses implicit expectations to find something that does fit. The way that this occurs in the model is that the two cross shapes are constructed in the lower layer (push), a situation is moved to in the higher layer that fits the new information (push), the expectations of the lower layer are changed (feedback) and the lower layer is able to construct from what is expected, resulting in the stable internal representation.

This demonstration provides confirmatory evidence for the first and third hypotheses through the connection between past experiences and movement to new situations. In the model, when the expectations held by the model do not serve, new expectations are used. Two questions that arise are: where do new expectations come from and how are new expectations tested? In this model all expectations were drawn from past experiences in the training phase. In answer to the second question, a technique was used of search through spreading activation within a layer.

3.2.2 Variation on the model

Two variations of the model were created to demonstrate different effects of constructive interpretation. In the first, a model was trained on similar shapes[2] in the same way as previously, Figure 9. The aim of the test phase was to observe a change of situation occurring in a gradually changing environment. In this test phase the network was presented with a representation resulting from the rules (e.g. three crosses) and moved from this through to another representation resulting from the rules (e.g. three squares) by a type of 'keyframing' over time, by first adding and then removing filled squares within the tartan grid, Figure 9.

[2] The only difference here is in the inclusion of a different kind of 'diagonal' cross

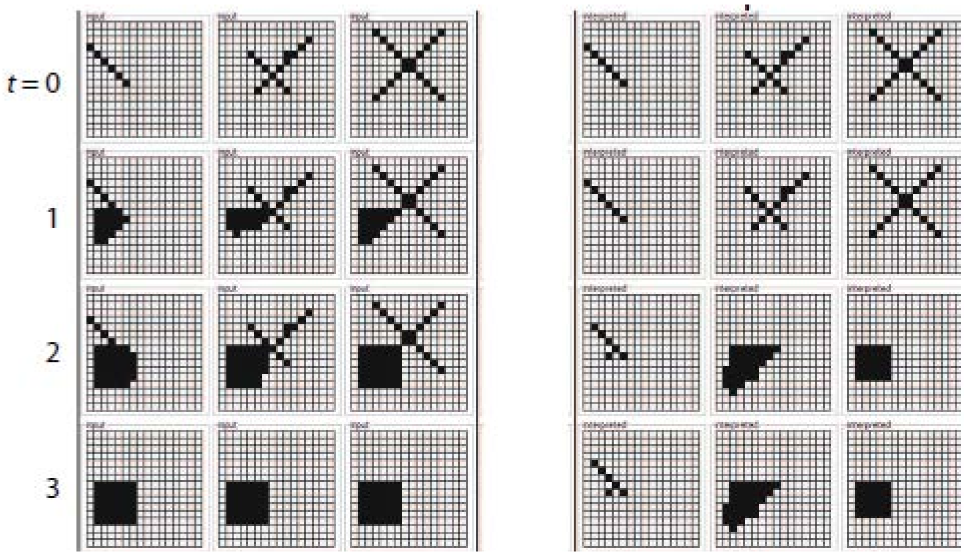

Figure 9 Four representations of a gradually changing environment (left block) and internal representation after interpreting (right block) as a time sequence

In Figure 9, at time *t=0* the model has a situation in which three crosses have been interpreted, and this continues to time *t=1* regardless that environment has changed further away from this interpretation. However, at time *t=2* the model can no longer construct the expected interpretation of three crosses; the situation changes based upon the data available to interpreting a cross, a triangle and a square.

Whilst this is a simple example, it represents the kind of phenomena seen during design activity where a designer can carry out activity within a situation for some time before, to an outside observer, there appears to be a large shift in the way that the design task is conceived – a change in the situation. At time *t=3* the environment changes further, and whilst it now appears to an outside observer as three squares, the network continues to interpret this as a cross, a triangle and a square, because in the world of the model what it is seeing is a fit with expectations.

This example shows a movement in which the situation is stable whilst the data changes, but once a threshold is reached in which the situation can no longer explain the data, it changes to account for this data. The phenomenon is evocative of the Sorites paradox, which observes that one grain of sand is not a heap and that by adding one grain of sand it does not become a heap and so induces that adding any number of grains will not make a heap. The analogy is useful, as the resolution of the paradox is related to our human conception of what it means for a number of grains of sand to be a ‘heap’. Similarly, in the demonstration, the small changes that are occurring in the environment become enough (in this case at time *t*=2) for the system to change the situation. The small changes to the environment can result in a stable interpretation that suddenly changes in a way that to an external observer is not commensurate to the small change in the environment.

Another variation serves to show that the same technique can be extended beyond working with shapes to show how the same phenomenon with letters creates a phenomenon that appears quite human, of seeing words to which we are habituated rather than a random assortment of letters. A similar training phase occurred in the setup for this demonstration, using letters instead of shapes[3], and the rules for constructing representations were to use only combinations of letters found in the English language[4]. In the example shown in Figure 10 the model sees the letters H,H,T as the word H,A,T. The way in which this constructed interpretation differs from traditional applications of AI algorithms (e.g. a typical SOM network) is that the interpretation is distributed over different layers, each attempting to construct from what is available and involves conceptual co-ordination over both space and time. Instead of a ‘one-shot’ approach in which either an interpretation is found or not, there is a dynamic convergence. In this case, the lower layer is not able to construct what it is expecting to find in the case of the letter A. However, the upper layer is able to construct what is it is expecting to find, a word that corresponds to experience. The feedback from this upper layer’s satisfaction of expectation feeds into the lower layer and the interpretation is settled upon – in this case from the top down.

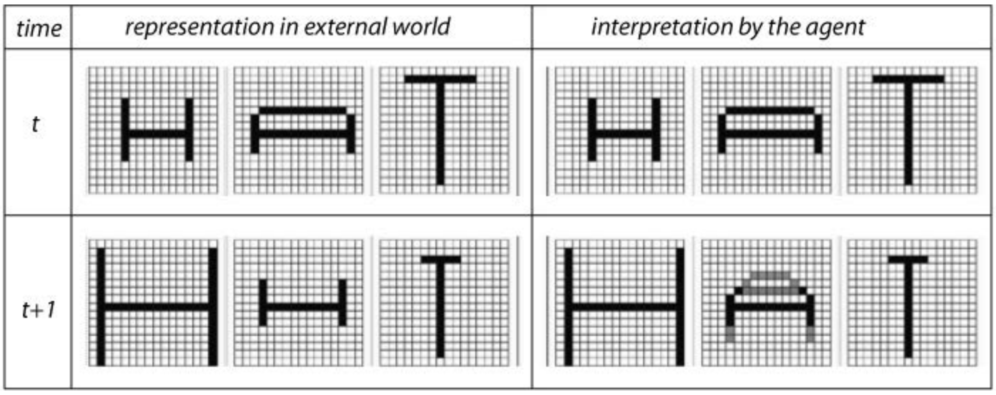

Figure 10 Conceptual co-ordination in space and time

[3] Algorithms were written for each of the letters A, E, H and T such that a letter of a random size was produced

[4] The words HAT, CAT, EAT, ATE

These demonstrations show some features of the model of constructive interpretation. The feedforward and feedback within each layer of the model serve to distribute this search for more suitable expectations. A suitable expectation does not mean 'a good fit for data' but rather 'a good fit for the situation'. The third demonstration with letters shows this, that the letter H had a good fit from perceptual experiences, but within the situation this was constructed as an A due to a higher layer finding an expectation that fit. This happened because the network was interpreting within a situation. In other words, the movement to a new set of expectations is always relative to where the model (or designer) currently is. In the models the guide for a change of expectations is from past experiences expressed through a search within each layer through spreading activation guided by similarity.

3.4 How does the situation change?

Designer are observed to move between different interpretations of a design medium. A question asked by observers of a designer is often 'why this novel interpretation and not some other?' The idea proposed here is that an interpretation is constructed from expectations and that there are expectations implicit in a situation that guide movements to new interpretations. This model shows how through constructive interpretation a system with experience can exhibit behaviour that looks to an external observer like unexpected discovery. This model provides evidentiary support for all three hypotheses.

A network similar to that described in the previous section was first trained on a set of 54 floor plans, Figure 11. These floor plans are a selection from the work of three prominent architects: Frank Lloyd Wright, Louis Khan and Andrea Palladio. They were used as there is existing work showing that there is a correlation between the perception of similarity in these sketches in humans and in neural networks (Jupp, 2005; Jupp & Gero, 2010). The model perceived each image by first using edge detection and a sharpening algorithm on the images and then feeding the resulting black and white pixels into the lower layer of the network as a sequence of 16x16 features. In the model a single floor plan becomes represented as a number of perceptual maps co-ordinated within a situation, Figure 12.

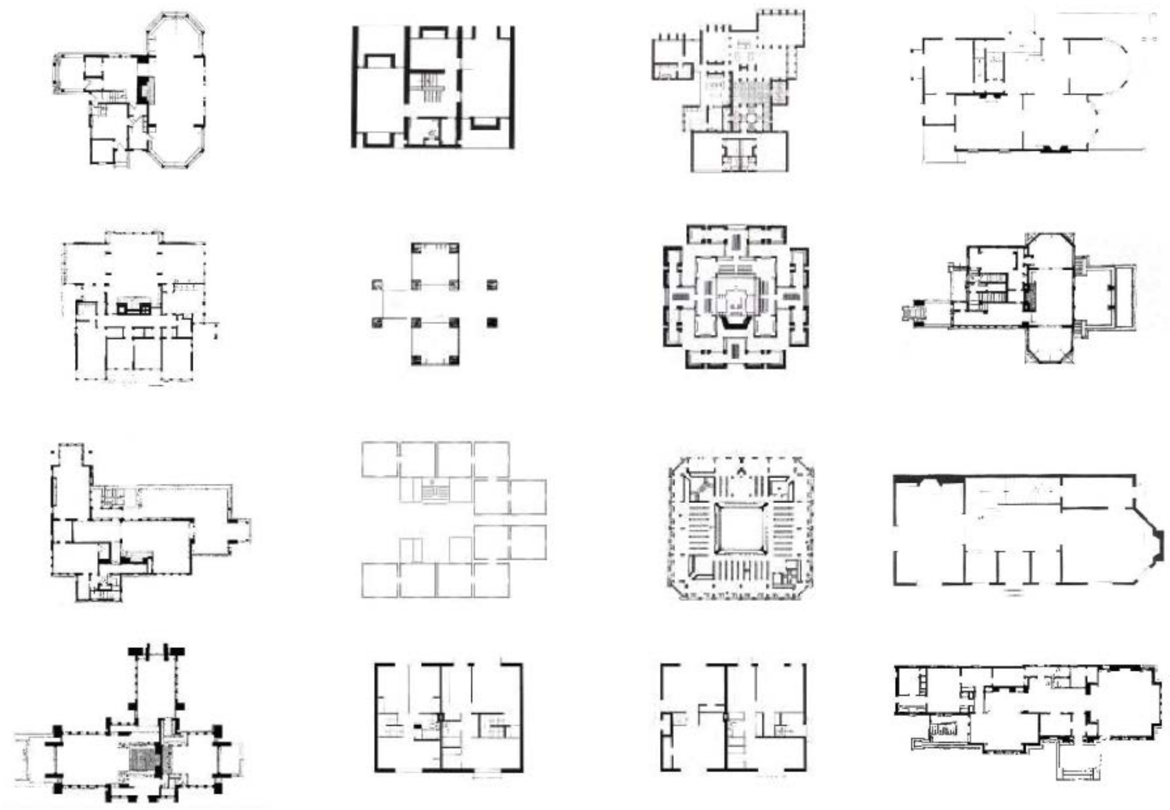

Figure 11 A sample of 16 floor plans from the full set of 54 plans by
Palladio, Lloyd Wright and Khan (source: (Jupp, 2005))

As a result of this training the set of these 54 floor plans come to be represented within the same two maps each at a different level of abstraction. At the lower level of the two networks, similarities can be measured across different experiences. For example, one of the perceptual maps from a Khan floor plan might be more perceptually similar (in terms of distance within the perceptual map) to a perceptual map from a Palladio floor plan than it is to another Khan floor plan.

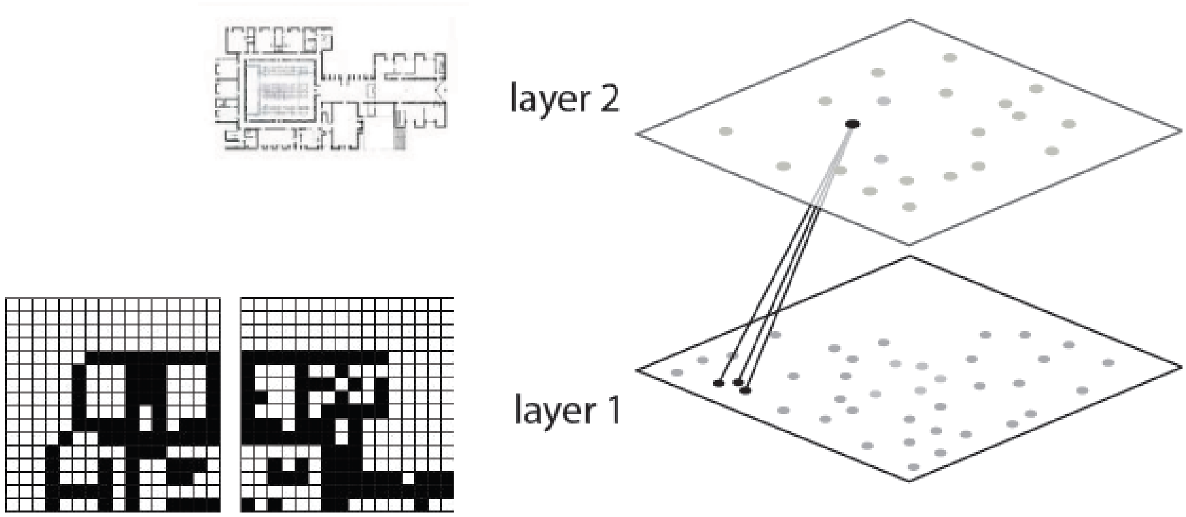

Figure 12 Representations of experience of floor plans within the network

This is one of a series of models that can be termed *generate-and-interpret* in which the actual generation is not so important as the effect that it has upon the situation. Rather than generate-and-test models which have a fitness function to use for testing, the aim of a generate-and-interpret model is to explore the different types of interpretation that can be computationally modelled – they aim to implement and test the way that a computational model explores its own knowledge.

The model begins activity with a set of expectations, taking one of the floor plans from experience and utilising the perceptual features associated with it for generating designs by randomly placing the perceptual features in the design medium – this is the starting point for the model. The system generates each step in the process by first interpreting the design medium, and then erasing it. It continues by randomly placing elements once again (the explicit expectations at the perceptual level) into the design medium. It uses the lower level maps that the system holds from its training in the different floor plans of well-known architects.

A snapshot of the model in progress is in Figure 13. The four images on the left-hand side show the four explicit perceptual expectations that the model is using in generating a design. The image at the top-centre shows the current state of the design external to the system in the design medium. Below this is in the centre is a representation of the floor plan from experience that is implicitly associated with these expectations. The right-hand side shows the perceptual features that have been used as a result of interpretation.

Figure 13 Interpreting using explicit concepts after random generation

Interpretation is done through a ‘sliding window’ that saccades across the image attempting to construct an interpretation from expectations and stopping when four interpreted percepts have been found. The heuristics for this movement are:

1. Attempt to construct from explicit perceptual expectations within the current window
2. Attempt to construct from implicit perceptual expectations within the current window
3. Move the window to the right; if the edge is found move the window down and start again from the left

Putting this together, the model follows iterations of:

1. commence with a set of explicit expectations
2. randomly generate a design using these expectations
3. interpret this design using its current expectations but potentially using implicit expectations
4. repeat from step 1.

In the example seen in Figure 13, the explicit expectations of percepts have all been found in interpreting – in other words, the situation after interpreting is no different to the situation before interpreting.

What happens when the model finds concepts that are implicitly expected is shown in Figure 14. Implicit expectations within the model are those that are similar (as defined by distance within conceptual space) to explicit expectations. What this means in practice is that as the model is generating and interpreting it creates novel interpretations for existing design elements. This changes the situation, as these implicit expectations are brought into the design situation. In the model, this was made into an iterative process by then drawing with these concepts that are the new explicit expectations. The result is that the model is exploring its own knowledge from experience in a series of steps. The guides for exploring its own knowledge are: (i) the conceptual similarity between explicit expectations and other expectations that exist from experience; and (ii) the relationships in the design medium that emerge.

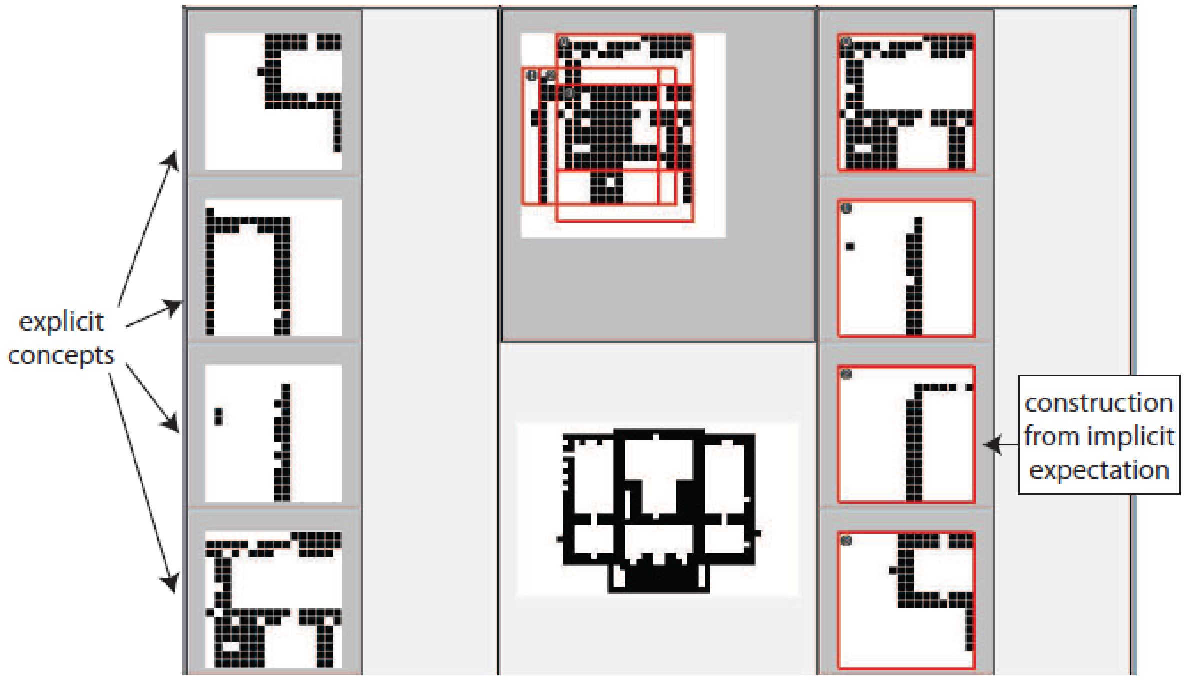

Figure 14 The system constructs an interpretation using two of implicit expectations and two explicit expectations

The effect of this is that the model works within its current understanding of the design problem until something triggers a change of situation, Figure 15.

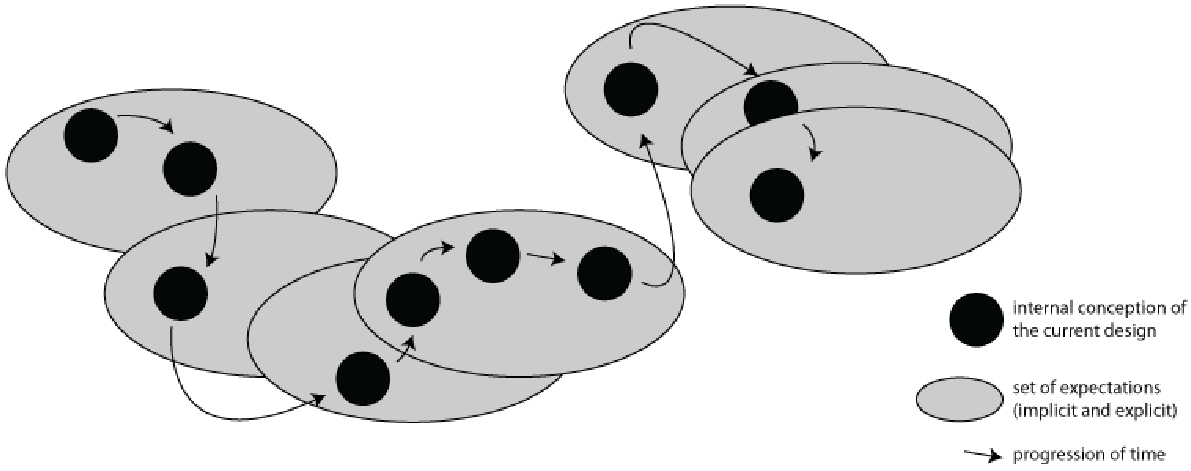

Figure 15 Grey ellipsoid is a situation that contains the set of implicit and explicit expectations. Black circle is the internal conception of the current design within this situation. Time progression is represented through black arrows of changing conceptions of design and changing situations.

In the model, making an interpretation can lead to the situation changing, with different expectations present. The contribution of the model is to show how the relationship between past experiences and these changes of situation. There is the potential for movement to new expectations within each layer of abstraction. This can lead to, for example, low level perceptual similarity serving as a cue for new higher level expectations. The model demonstrates how the hypotheses that emerge from this model of interpretation relate to phenomena observed in designers.

3.5 Discussion from the models

Two questions that arise are: where do new expectations come from and how are new expectations tested?

The work can be seen as an example of models that generate-and-interpret, as distinct from those that generate-and-test. This type of model makes a contribution in the way that it navigates its own knowledge through interpretation. In the models described here the design actions were random and unconsidered, but the models are interesting because of the implementation of interpretation that resulted in the model exploring its own knowledge. In the model described in Section 3.4 it was shown that it could commence working with notions drawn from sketches by Louis Khan, and could “see” things within the (randomly produced) designs that caused it to bring notions from Frank Lloyd Wright sketches into the design situation.

There are many ways that interpretation could be implemented as inspired by recent ideas from cognition. Generate-and-interpret models facilitate testing these different implementations.

The key observations from the demonstrations produced from this model of situated design are:

1. Expectations exist at different levels of abstraction
2. Interpretation involves three activities: (i) constructing from existing expectations using data available; (ii) changing expectations when existing expectations do not work; and (iii) distributing this across multiple layers of abstraction through feedforward of data and feedback of expectations
3. Change of expectations is relative to the current expectations; an analogy is that past experiences are a map, current expectations are a location and change of expectations is movement within this landscape. Whilst this sounds like spreading activation, it functions differently due to the distribution of expectations across different layers of abstraction and the links between them.
4. Stability exists in a slowly changing environment, followed by a sudden jump to a different situation. This maps well onto the kind of behaviour that is observed in designers making unexpected discoveries

3.5.1 Potential of the models

The implementation described here makes use of two linked SOM maps, but the ideas could be implemented in many other ways. Similar results would be expected with ART networks (Carpenter & Grossberg, 1988) and potentially hierarchical ART networks that could be adapted to fit (Carpenter & Grossberg, 1990; Tscherepanow, 2010). Adapting existing hierarchical unsupervised learning systems such as deep belief networks (Hinton, Osindero, & Teh, 2006; Lee, Grosse, Ranganath, & Ng, 2009) to hold expectations in a way that fits the notions of situatedness described here, is a potential avenue of further research.

Whilst the models described use single-thread programming, future models can explore what happens when this process occurs in parallel within each layer to set up a dynamic system.

The implementations appear to fulfil some of the expectations of cognitive studies of design described in Section 2. An empirical study of a model could be facilitated using these techniques comparing the two explicitly and finding generate-and-test implementations for interpretation that better suit what we observe in human designers.

4 Conclusions

This paper has articulated a type of knowledge representation appropriate for situated design. It has demonstrated one way of computationally implementing this knowledge representation and has provided support for hypotheses about design that result from this.

The work has examined how situations change during design activity. In this paper the expectations held by a designer have been linked to the process of interpretation, and a change of situation linked to a change of expectations. These models demonstrate the way that interpretation uses expectations resulting in the type of phenomena observed in designers bringing knowledge into the design situation and changing their ideas about the activity in which they are engaged.

It has been proposed that changes to the situation happen at multiple of abstraction within such a model of a designer; and that when expectations are not met, the implicit conceptual similarity guides the movement to new expectations. This can be observed in the results from the computational implementation, which provide support for three hypotheses about design:

1. It was hypothesised that attempts to construct an interpretation using expectations can lead to a change in the situation. This was observed within the models where the process of attempting to construct an interpretation from expectations, when these were not met, led to a change of expectations. This resulting change of situation, to an external observer, looks like the phenomenon of changing situations in a designer
2. Small changes in a source can lead to large changes in the internal representation. It was observed that small changes in the stimulus outside the model can result in a large change in the situation. This was observed most clearly in the model of a gradually changing environment that maintains the situation until it is no longer useful, at which point the situation can change quite dramatically.
3. Changes to the situation have their origin in the experiences of the design agent, manifested as connections between knowledge. When the situation changes, it

changes in ways that are related to the first person knowledge of the design agent from past experiences. In the model one heuristic for a way that this could occur has been expressed, as spreading activation at different layers of abstraction.

The primary contribution of this work is in utilising concepts from situated cognition to model the way that designers explore their own knowledge through interpretation in a way that matches observed behaviour

Acknowledgements

This work has been supported by an Australian Postgraduate Award scholarship, by the Australian Research Council under grant no. DP 0559885, and by the US National Science Foundation under grant no. SBE-0915482. Writing support has been provided by the Pontificia Universidad Católica de Chile through the MECESUP PUC0611 project. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the Australian Research Council or of the US National Science Foundation.

References

Barsalou, L. W. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences, 22*, 577-660.

Barsalou, L. W. (2005). Abstraction as dynamic interpretation in perceptual symbol systems. *Building object categories*, 389-431.

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

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1521), 1281-1289.

Begg, I., Anas, A., & Farinacci, S. (1992). Dissociation of processes in belief: source recollection, statement familiarity, and the illusion of truth. *Journal of Experimental Psychology, 121*, 446-458.

Beyer, O., Cimiano, P., & Griffiths, S. (2012). *Towards action representation within the framework of conceptual spaces: Preliminary results.* Paper presented at the Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence.

Bilda, Z., Gero, J. S., & Purcell, T. (2006). To sketch or not to sketch? That is the question. *Design Studies, 27*(5), 587-613.

Carpenter, G. A., & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer, 21*(3), 77-88.

Carpenter, G. A., & Grossberg, S. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural networks, 3*(2), 129-152.
Clancey, W. (1997). *Situated Cognition: On Human Knowledge and Computer Representations*: Cambridge University Press.
Clancey, W. (1999). *Conceptual Coordination: How the Mind Orders Experience in Time*: LEA, Inc.
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* (Vol. Baltimore, pp. 620-622). Baltimore: Lippincott Williams and Wilkins.
Cross, N. (1982). Designerly ways of knowing. *Design Studies, 3*(4), 221-227. doi: 10.1016/0142-694x(82)90040-0
Dittenbach, M., Merkl, D., & Rauber, A. (2000, 2000). *The growing hierarchical self-organizing map.* Paper presented at the Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on.
Fish, J., & Scrivener, S. (1990). Effects of environmental context on recognition memory in an unusual environment. *Perceptual and Motor Skills, 63*, 1047-1050.
Fodor, J. A. (1975). *The Language of Though*: Harvard University 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.
Gero, J. S. (1990). Design prototypes: A knowledge representation schema for design. *AI Magazine, 11*(4), 26-36.
Gero, J. S. (1998). Conceptual designing as a sequence of situated acts. *Artificial intelligence in structural engineering*, 165-177.
Gero, J. S. (2007). Situated design computing: Principles. In B. Topping (Ed.), *Civil Engineering Computations: Tools and Techniques* (pp. 25-35). Stirlingshire: Saxe-Coburg Publications.
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.
Gero, J. S., & Smith, G. (2009). Context, situations and design agents. *Knowledge-Based Systems, 22*, 600-609.
Gombrich, E. H. (1966). *Norm and Form: Studies in the Art of the Renaissance*. Oxford: Phaidon.
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

Gross, M., & Do, E. Y.-L. (2000). Drawing on the back of an envelope: a framework for interacting with application programs by freehand drawing. *Computers and Graphics, 24*(6), 835-849.
Harnad, S. (1999). The symbol grounding problem. *Physica D, 42*, 335-346.
Hawkins, J. (2005). *On Intelligence*. New York: Holt Paperbacks.
Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. *Neural computation, 18*(7), 1527-1554.
Jupp, J. (2005). *Diagrammatic reasoning in design: computational and cognitive studies in similarity assessment.* PhD, The University of Sydney, Sydney.
Jupp, J., & Gero, J. (2010). Let's Look at Style: Visual and Spatial Representation and Reasoning in Design. In S. Argamon, K. Burns & S. Dubnov (Eds.), *The Structure of Style* (pp. 159-195): Springer Berlin Heidelberg.
Karmiloff-Smith, A. (1994). Beyond modularity: A developmental perpective on cognitive science. *Behavioral and Brain Sciences, 17*, 693-745.
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
Lee, H., Grosse, R., Ranganath, R., & Ng, A. Y. (2009). *Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations.* Paper presented at the Proceedings of the 26th Annual International Conference on Machine Learning.
Menezes, A., & Lawson, B. (2006). How designers perceive sketches. *Design Studies, 27*(5), 571-585.
Minsky, M. (1974). A Framework for Representing Knowledge.
Montare, A. (1994). Knowledge acquired from learning: new evidence of hierarchical conceptualization. *Perceptual and Motor Skills, 79*(2), 975-993. doi: 10.2466/pms.1994.79.2.975
Mountcastle, V. (1997). The columnar organization of the neocortex. *Brain, 120*, 701-722.
Murphy, G. L. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.
Newell, A., & Simon, H. A. (1961). *GPS, a program that simulates human thought*: Defense Technical Information Center.
Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(1), 54.
Parmet, A., & Gillingham, K. (2002). Spatial orientation. In R. Dehart & J. Davis (Eds.), *Fundamentals of Aerospace Medicine* (pp. 184-244). Philadelphia: Williams and Wilkins.
Peng, W., & Gero, J. (2006). Concept formation in a design optimization tool. *Innovations in Design & Decision Support Systems in Architecture and Urban Planning*, 293-308.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., McRae, K., & Spivey, M. J. (2011). The mechanics of embodiment: a dialog on embodiment and computational modeling. *Frontiers in psychology, 2*.
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 Ed.* (pp. 665-666). Baltimore MD: Lippincott Williams and Wilkins.
Piaget, J. (1954). *The Construction of Reality in the Child*: Basic Books.
Pylyshyn. (1984). *Computation and Cognition*: Harvard University Press.
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.
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.
Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale: Lawrence Erlbaum Associates.
Schacter, D. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(3), 501-518.
Schon, D., & Wiggins, G. (1992). Kinds of seeing and their functions in designing. *Design Studies, 13*(2), 135-156. doi: citeulike-article-id:8497732
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.
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.
Thoreau, H. D. (1851/1993). *A Year in Thoreau's Journal: 1851*. New York: Penguin Books.
Tscherepanow, M. (2010). TopoART: A topology learning hierarchical ART network *Artificial Neural Networks–ICANN 2010* (pp. 157-167): Springer.
Wright, P. C., & Monk, A. F. (1985). Evaluation for design. In A. Sutcliffe & L. Macualay (Eds.), *People and Computers V* (pp. 345-358). Cambridge: Cambridge University Press.
Zusne, L. (1970). *Visual Perception of Form*. New York: Academic Press.