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 a model of the way that designers move between situations when interpreting during design activity. Three hypotheses are presented that arise from this model: that designers change their situation during interpretation, that small changes in a source can lead to large changes in the representation, and that changes to the situation have their origins in the experience of the designer. The paper demonstrates how this internal movement between situations can be computationally implemented using three examples. The systems implemented demonstrate the way that interpretation can lead to changes in the situation and present an example of how the changes to a designer's situation can be guided by past experiences.

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. It has long been recognised that the expectations held by a designer (the way that they 'see') have an effect upon their future design moves (Schon & Wiggins, 1992). This work is concerned with the question of when do a designer's expectations change and why do they change to one set of expectations and not another? It is an enquiry into the relationship between experience and expectation as observed in a design conversation.

Designers' notions of what they are doing change during design activity, and through think-aloud studies and video recordings it is possible to observe, to some degree, what their notions are changing from and to (Fish & Scrivener, 1990; Gross & Do, 2000; Suwa & Tversky, 1997). The motivation for this work is to move beyond observation and to posit the cognitive processes that relate knowledge from experience to the changing conception of a design task, with the aim of making computational models of design more 'designerly' (Cross, 1982).

To place the focus upon the cognitive processes of the designer, design will be conceived here as a sequence of actions by a designer, where these actions are distinguishable activities over time that take place within a situation (Gero, 1998). The term *situation* refers here to a designer's internal notion of the world at a particular time, their co-ordinated expectations (Clancey, 1997). 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]. The situation refers to the parts of grounded knowledge (Barsalou, 2007) that come from experience within the world and that are being used to understand the world at this current moment and in this current experience.

The paper first describes a model linking the situation of a designer (and their expectations within this situation) to the cognitive process of interpretation – something that occurs frequently during the observed seeing-moving-seeing activities of design. It then uses computational models to demonstrate how these notions can be used to make models of design more designerly, moving from one situation to another in a way that is inspired by empirical studies of designers.

1 Background

1.1 First person knowledge

Knowledge can be grounded in experience leading to knowledge that can be described as *first person knowledge*. Although 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 (Chisholm, 1982). 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"

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

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.

1.2 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 (Pylyshyn, 1977) towards interpretation as situated conceptualisation (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, the situation changes (Gero & Kannengiesser, 2004). The motivation for adopting this perspective is to make use of the cognitive processes of changing interpretations of the world that appear to occur during design in modelling designing. Existing models of interpretation in design have focussed upon analogical transfer (Goel, 1997; Yaner & Goel, 2008).The focus in the situated cognition approach is upon the relationships between experience, expectation and interpretation.

1.2.1 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, Gero, & Purcell, 2000). When designers look at their own work they are capable of inventing new design issues and intentions, 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 intention for the design.

An example of first person knowledge and the movement from one set of expectations to another can be seen in this example:

> "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)

In this description, the designer has commenced an activity within one situation (A), 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 of the designer changes (B), bringing in a new goal to find a way of having the water and the artworks co-exist. The question addressed by this paper is about how this movement of situation from A (within which no solution is possible) to B (in which a solution is possible) occurs for the designer. A system with the required knowledge might be able to use a symbolic AI approach (typified by the subgoaling strategy of Newell (1994)) but would not arrive at the same class of solution as the designer in the example unless it had unique strategic knowledge about how to connect these two disparate domains. Even then the way an AI system might arrive at a solution would not be called designerly.

In the example, 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. In this paper the hypothesis is adopted, following Schon and Wiggins, that the act of interpreting can cause the situation to change. The rest of this paper is concerned with describing how this may be explained and presenting computational models of this movement.

2 How designers move between situations during design activity

The remainder of this paper is concerned with describing a way that computational models could be made more designerly in their movement between situations. This claim is based upon specific conceptions of what is meant by: (i) knowledge representation; (ii) expectations; and (iii) interpretation. The contributions of the paper are the model of interpretation and three hypotheses about the way that designers move between situations. The three hypotheses are:

1. the process of ‘constructing’ an interpretation can lead to a change in the designer’s situation;
2. small changes in the source (of an interpretation) can lead to large changes in the internal representation; and
3. changes to the situation have their origins in the experiences of the designer.

2.1 Knowledge representation

An attempt to model interpretation computationally requires a computational model of knowledge. 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). This feature can be observed within knowledge as described as a *Perceptual Symbol System* (PSS; Barsalou, 1999, 2005). 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). In a PSS, 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 PSS provides a paradigm for knowledge representation, and this is instantiated in our models using the *conceptual spaces* framework of Gärdenfors (2000). The conceptual spaces model has the advantage of being demonstrably useful for computation (Beyer, Cimiano, & Griffiths, 2012; Gärdenfors, 2000) and meets the requirements for a situated design agent outlined by Gero and Fujii (2000) once the layer of situation is added to this framework. A brief description of this framework with the required modification is presented.

2.1.1 Conceptual spaces (with the addition of situations)

Abstraction is critical for representing knowledge in a PSS (Barsalou, 2005). In this paper we will refer to four levels of abstraction as a useful delineation, following on from Gärdenfors (2000) in describing sensation, perception and conception and Clancey (1999) in adding situations to co-ordinate 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* (Montare, 1994). 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.

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.

Situations can be thought of as a convergence zone for these concepts – the memory of the co-ordination of concepts (Barsalou, 1999; Clancey, 1997, 1999). Concepts that are utilised at the same time within an agent are a part of the same situation. The framework of Gärdenfors (2000) is extended here 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.

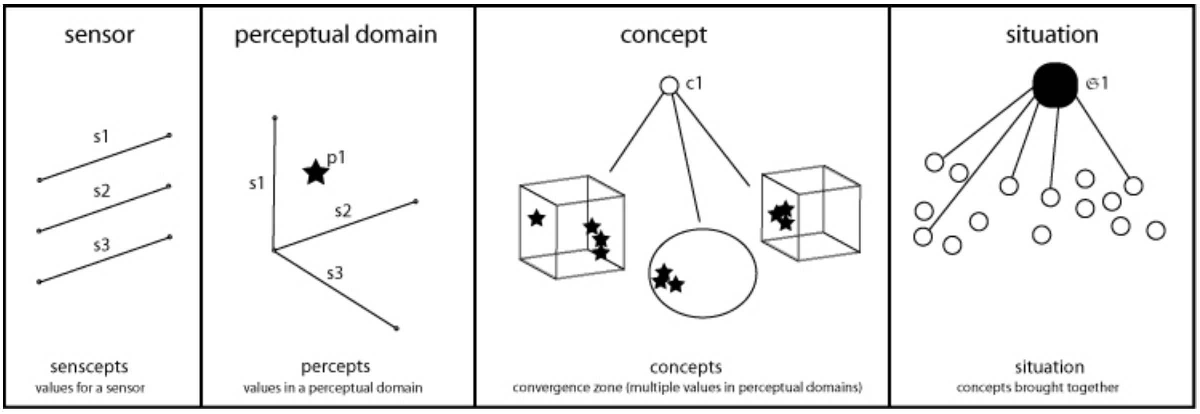

Figure 1 Levels in a hierarchy of knowledge organisation for design

2.2 Expectations

A designer holds knowledge about the world from their experience. At a certain moment in time, some of this knowledge is explicitly being utilised as the explicit expectations of a designer. Asking the question "What are you doing now?" (Clancey, 1999) and think-aloud studies (Wright & Monk, 1985) are ways to get a sense of these expectations.

This explicit knowledge is a part of the situation of the designer. In addition the designer holds *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 connections which exist but of which the agent is not aware. *Implicit memory* is "when performance on a task is facilitated in the absence of conscious recollection" (Graf & Schacter, 1985).

The effect of implicit memory can be observed in studies of priming, 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). Priming phenomena make use of connections from past experiences to improve the speed of recognition.

2.3 Interpretation within a situation

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. Following Piaget (1954), interpretation is concerned with reconciling the two opposing goals of: (i) maintaining a stable notion of the world; and (ii) changing the picture of the world when the world changes. These two goals can be described following Gero and Kannengeisser (2004) as: (i) a *pull* from expectations in each layer of the knowledge hierarchy attempting to construct what is expected from the information arriving about the world; and (ii) a *push* of data from the external world flowing up the hierarchy of abstraction causing expectations to be changed.

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 on observing 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.

2.3.1 A model of interpretation as a constructive activity

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. The term *source* is used to refer to the object of interpretation because it recognises that from any source it is possible to pull many different kinds of interpreted data. 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 one part of interpretation that allows for the unexpected to be recognised and allows for new knowledge to be learned.

It is through push that the first hypothesis arises, which holds that push from source data can change the expectations of an agent during interpretation. When the current expectations of an agent are not able to account for the data from the source, a change of expectations are expected to occur.

A layer in a system of interpretation is shown in Figure 2, in which the curved arrow represents an attempt to reconcile data from the source with expectations held prior to interpretation. All layers in this proposed system utilise the process of interpretation, but at a different level of abstraction.

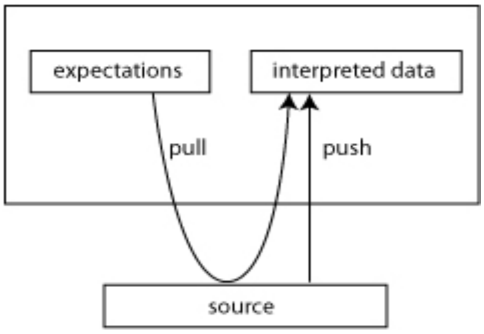

Figure 2 Interpretation within a layer as push from a source and pull from expectations, represented as a curved arrow attempting to create the expectations using data from the source

An example of pull is used to help understand the description of the pull arrow used in Figure 2. Participants in an experiment were played an audio track, consisting of a continuous musical note interrupted by white noise, followed by the resumption of the original note. People in this experiment 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 from the white noise the continuation of the note that it expects to be finding.

Figure 3 shows a dynamic system constructed from multiple self-similar layers of the type seen in Figure 2, inspired by the work of Mountcastle (1997). Each layer attempts to construct what it expects to find from what is available in the layer below. The layer makes data available to the layer above, and feeds back information to the layer below to alter its expectations. Another way to describe this is that interpreted data flows up and changed expectations flow down.

This is the basis for the second hypothesis which is that the system for interpretation seen in Figure 3 is a complex system, in which small departures from expectations can lead to large changes in the world view of the agent. It is difficult to predict the changes that will occur in the system when one layer is changed, as either a small change in the topmost layer can lead to large changes in the layers below or a small change in a lower layer can result in all layers changing.

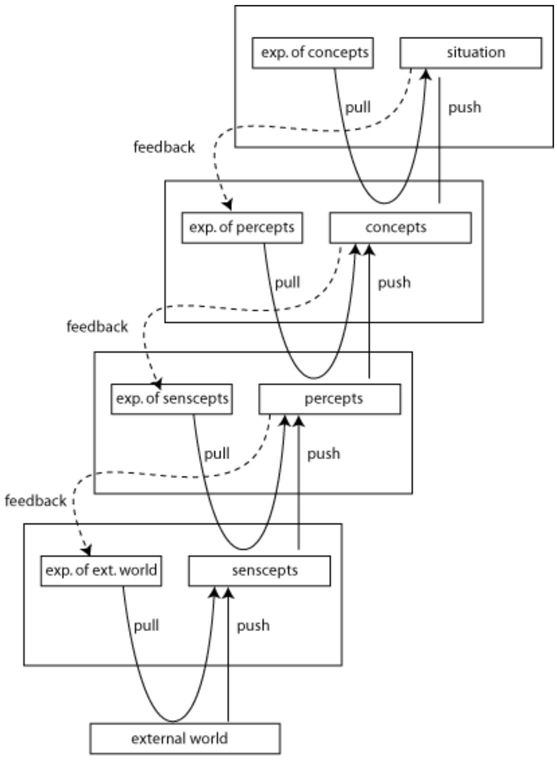

Figure 3 The dynamic system resulting from multiple layers with the presence of feedback from the interpreted data of one layer to the expectations of the layer below

The third hypothesis has its origins in the observation that the 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 suggestion is that the design conversation with a medium is a way of the designer navigating their knowledge from experience.

3 Modelling movement between situations during design activity

These three hypotheses are further explored through computational models of interpretation. The models use the conceptual spaces knowledge representation described in Section 2 to co-ordinate their notions about the world in a situated way, and perform constructive interpretation through push and pull. A description of these models serves to demonstrate the computability of the approach and to give preliminary evidence to support the hypothesis that the type of interpretation described here results in systems that change their expectations in a designerly way.

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) and for experimenting with feedback and feedforward within the resulting hierarchy.

The models described use two linked SOMs to create a model of interpretation with two layers of abstraction. Figure 4 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. 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 5. Step 1 in Figure 5 is an attempt to pull from expectations in layer 1. If pull cannot occur, then step 2 is to 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.

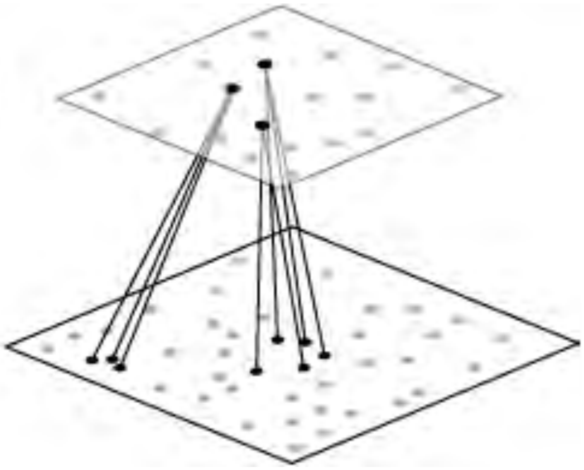

Figure 4 Linked SOMs representing abstraction from data.

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.

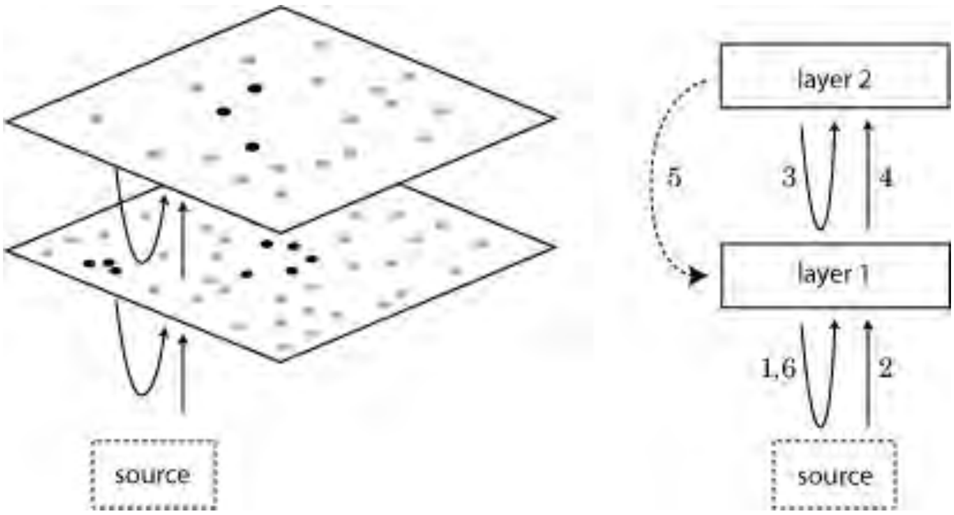

Figure 5 Push-pull in a two layer system with numbered steps: 1 pull from expectations in layer 1; 2 push from the source into layer 1; 3 pull from the expectations of layer 2; 4 push from layer 1 into layer 2; 5 an update of layer 1 expectations by layer 2; and 6 interpretation through pull from layer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

3.1.1 Describing the implementation

Two SOMs are implemented to represent the lower layer of abstraction SOM^1 and the 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.

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

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

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. 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 the BMU. In the models described here expectations are implemented as a subset of all nodes.

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

The similarity in the models is defined by the Nosofsky distance (Nosofsky, 1988) between the input and best matching unit $d_N(x, m_{u,v}) = e^{-c*(d_E)^2}$. 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.

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

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 Implementations to demonstrate the three hypotheses

Three implementations of this approach serve to model aspects of the three hypotheses. Two implementations highlight specific consequences of constructive interpretation, whilst the third implementation provides a model of how this looks in a synthesised design conversation.

3.2.1 Changing situations in a model of constructive interpretation

Two linked self-organising maps as seen in Figure 5 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 6 gives an example of a representation of three squares produced in this way. The model makes use of expectations that are both spatial and temporal in nature. 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. 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.

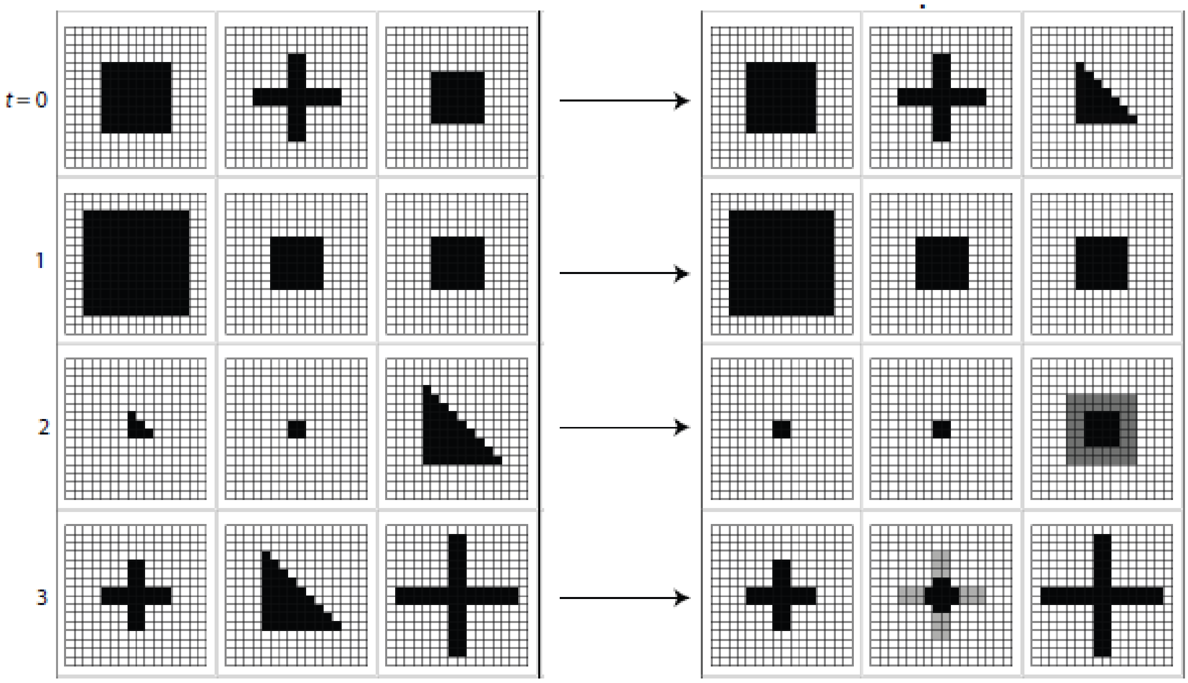

Figure 6 Four representations (left block) and internal representation after interpreting (right block) as a time sequence (the light grey is an artefact of the abstraction that occurs during training of the SOM).

In the test phase of the model following training, a series of three random shapes were produced to create representations that may or may not obey the rules presented during training. For example, two squares and a triangle was a possible representation in the test phase. Constructive interpretation occurred in the test phase of the model through:

(i) **Temporal explicit expectations:** The trained network held 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 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.

The motivation for this model is to provide confirmatory evidence for the first and third hypotheses. In this model it can be seen that the process of 'constructing' an interpretation can lead to a change in the situation. The active nodes in the layers of the model represent explicit expectations about what will be observed. In some cases, what it is expecting to see will be used in constructing an interpretation, and have the model seeing something different to what is observed by a human observer. In other cases the model will not be able to construct from these active nodes, and in changing active nodes can be said to have its expectations changed by interpretation, moving it to a different situation. In this case of changed expectations, the third hypothesis gives an understanding of the relationship between past experiences and the model moves to one situation and not another.

An example of the results produced by the model is shown in Figure 6. The hypothesised 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 due to its expectations. 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 because it can only work with what it is presented, the higher layer does the same. In this case the temporal expectation has been used to construct the internal representation. The case at time $t=3$ is an example of change of situation occurring through interpretation. The temporal expectations have not been useful as 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 example can be used to describe the third hypothesis. The model holds a set of explicit expectations in each of the two layers, modelled as active nodes. The nodes that are proximal to these nodes, at each layer of abstraction, represent implicit expectations. They are implicit because when expectations change, these are the first expectations to be used without being liminal . And the reason that these nodes are proximal is that during training the association was reinforced. In this way a relationship can be established between past experiences, the current situation and why a move occurred to another particular situation.

3.2.2 Variation on the model

A variation on this model serves to explore the second hypothesis; that small changes in the sources can lead to large changes in the interpretation. The model was trained in the same way on similar shapes[2], Figure 7. The difference in this model was that that rather than a different unrelated representation at each point in time, the model had a gradually changing environment. The environment changed over time, from a representation that followed the rules (e.g. three crosses) and then moved from this through to another representation resulting from the rules (e.g. three squares). This movement occurred over time by a type of 'keyframing', first adding and then removing filled squares within the tartan grid, Figure 7.

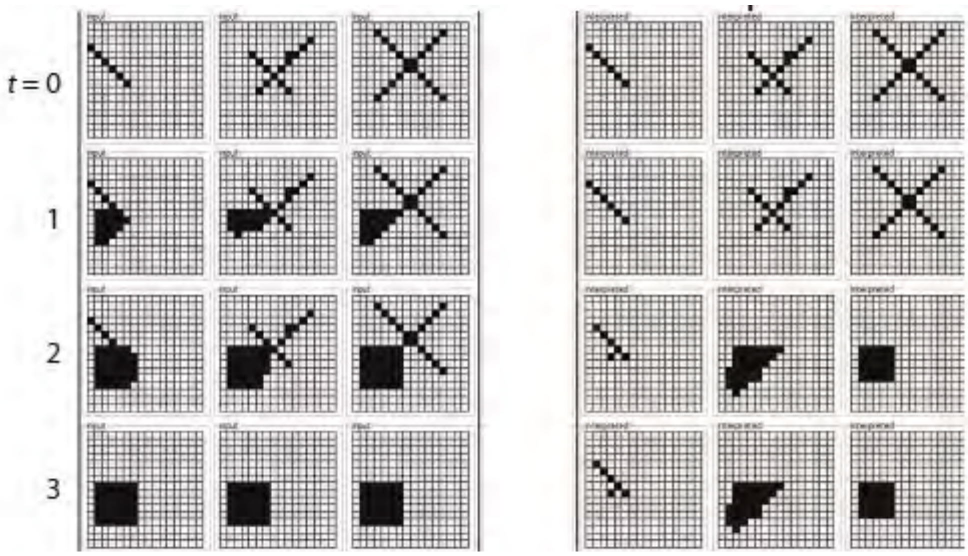

Figure 7 Four representations of a gradually changing environment (left block) and internal representation after interpreting (right block) as a time sequence

In Figure 7, at time *t=0* the model has interpreted the source as three crosses. This interpretation is unchanged at time *t=1* despite the observation that environment has changed during this time, as interpretation is still able to construct what it is expecting from what is available. However, at time *t=2* the model can no longer construct the expected interpretation

---

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

of three crosses. A small change in the environment has led to the situation changing based upon the data available. Now the source is interpreted as a cross, a triangle and a square.

This is an example of the second hypothesis, and 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 with a change in 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.

A further variation demonstrates an extension of the approach beyond working with shapes. The same phenomenon can be observed in an environment composed of 'letters' to create the phenomenon of seeing words to which we are habituated. A similar training phase occurred in the setup for this demonstration, using letters as a special type of shape[3]. The rules for constructing representations were to use solely combinations of letters found in the English language[4]. In the example shown in Figure 8 the model sees the letters HHT as the word HAT. This same result can be achieved with a straightforward application of an AI algorithm (e.g. a typical SOM network). The suggestion is that the way in which the model arrives at an interpretation, by distributing the interpretation over layers of abstraction with each attempting to construct from what is available, is more designerly. It involves conceptual co-ordination over both space and time and instead of a 'one-shot' approach in which either an interpretation is found or not, there is a dynamic convergence. In the example 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.

---

[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

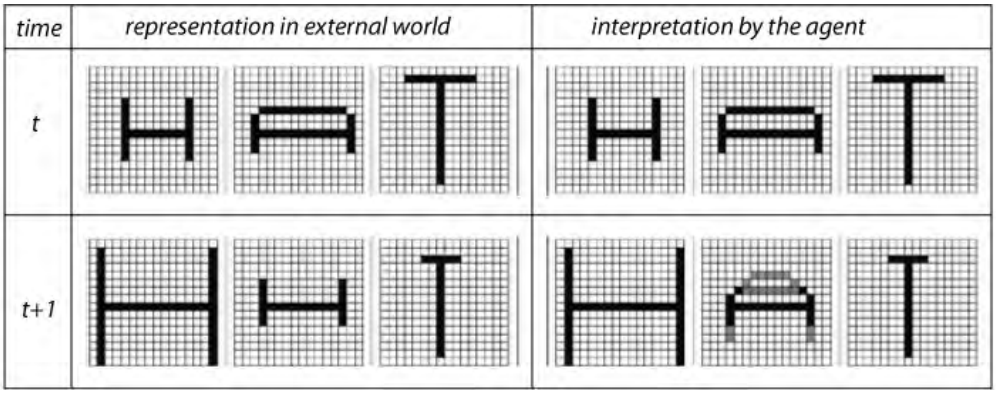

Figure 8 Conceptual co-ordination in space and time

This simple model of constructive interpretation and its variations provides a demonstration of the hypotheses proposed in section 2. In the models, interpretation occurs in a situation. The expectations in this situation guide the construction of an internal representation, and also implicitly suggest the next situation if expectations change during interpreting. The model of constructive interpretation is one of dynamic convergence distributed across layers of abstraction, with feedforward of data and feedback of expectations.

3.4 A model of constructive interpretation in the design conversation

A further model demonstrates how these hypotheses relate to our understanding of designers engaged in a design conversation with a medium. As observed by Schön and Wiggins and in studies of unexpected discovery, designers 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 three hypotheses of constructive interpretation are modelled in a system engaged in designerly behaviour of 'moving' and 'seeing'.

A network was first trained on a set of 54 floor plans, Figure 9. 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 system 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 10.

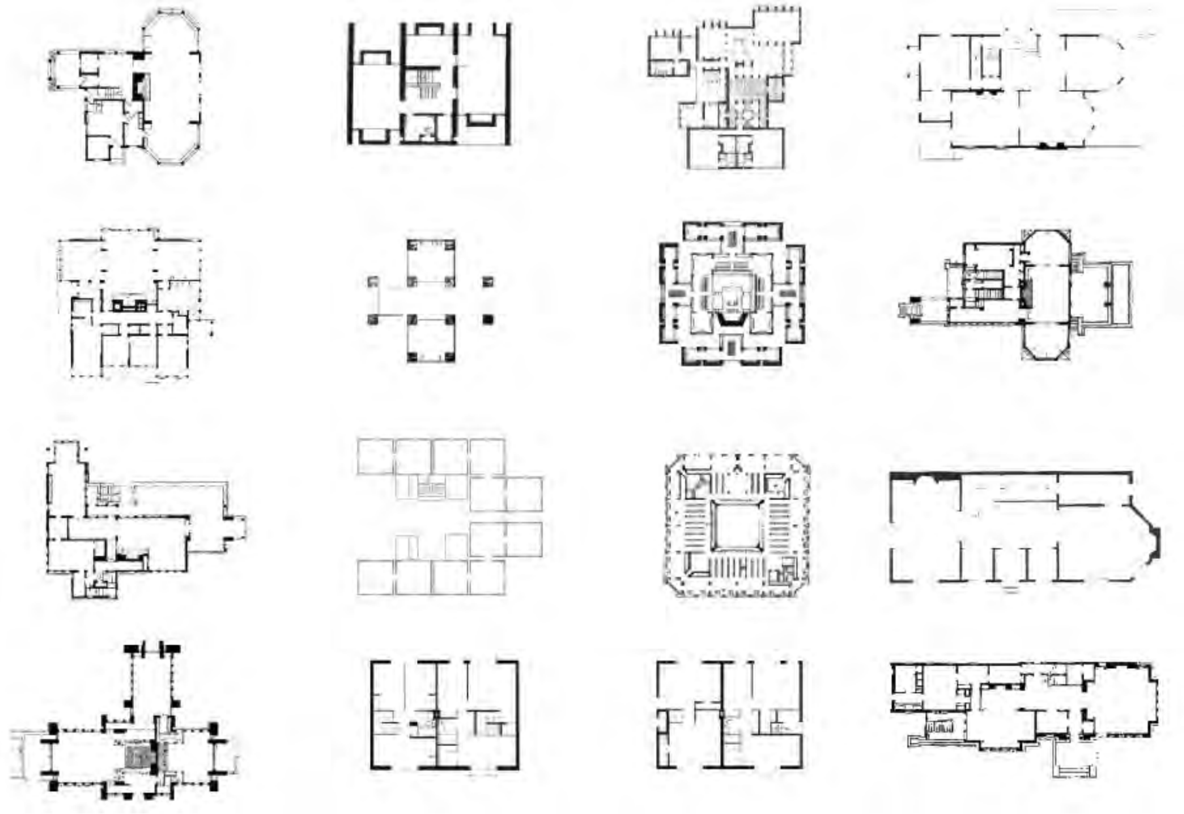

Figure 9 A sample of 16 floor plans from the full set of 54 plans by Palladio, Lloyd Wright and Khan (source: (Jupp, 2005))

As a result of this training the set of these 54 floor plans comes 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.

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.

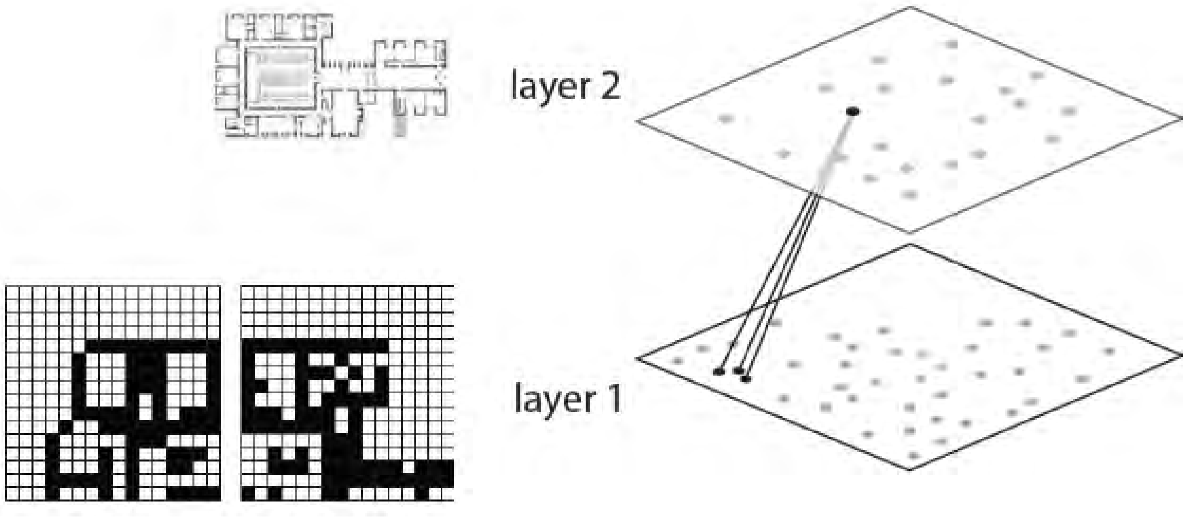

Figure 10 Representations of experience of floor plans within the network

A snapshot of the model in progress is in Figure 11. 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 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 11 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; and
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; and
4. repeat from step 1.

In the example seen in Figure 11, 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 12. 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 as 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.

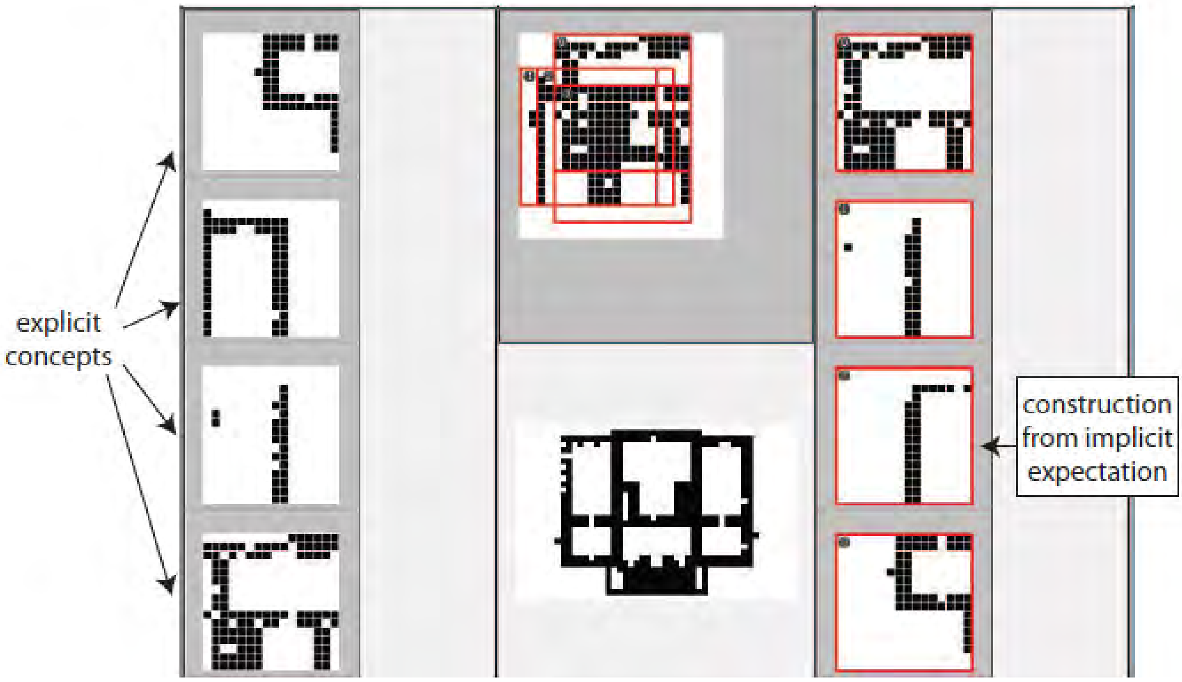

Figure 12 The system constructs an interpretation using two of implicit expectations and two explicit expectations

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 13.

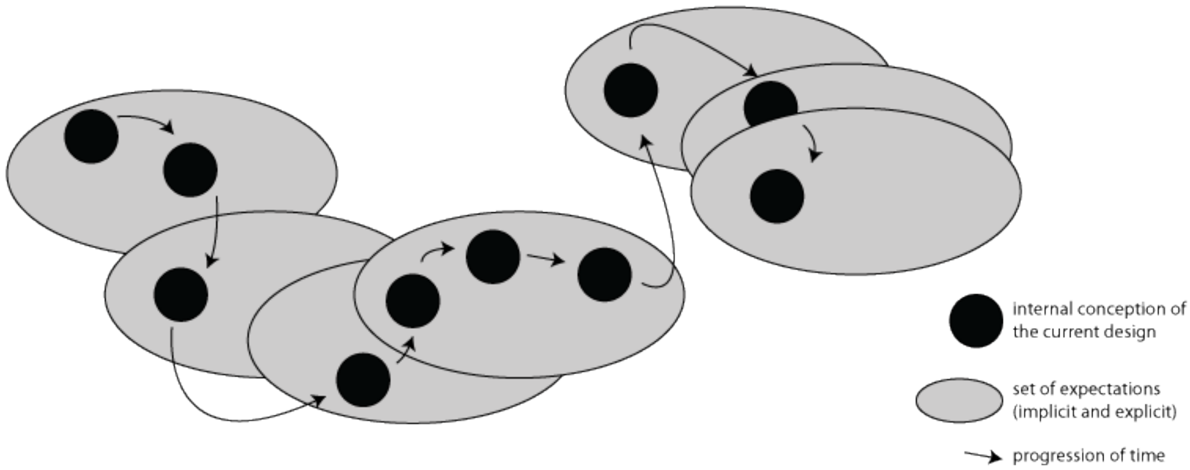

Figure 13 Each grey ellipsoid is a situation that contains the set of implicit and explicit expectations. Each black circle is the internal conception of the current design within this situation. Time progression is represented through black arrows of changing conceptions of design and changing situations.

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, for example, to 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 may come to be related to phenomena observed in designers.

3.5 Discussion from the models

In Section 2 a model of interpretation that fits with a description of situated design was presented. Selected results have been presented to show how aspects of such a system can be implemented. The critical features of the models as they relate to the three hypotheses can be summarised as follows.

- Expectations exist at different levels of abstraction.
- 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. This relates to the first hypothesis, that interpretation can change a designer's situation.
- Through push (from data) and pull (from expectations) there is a balance between a stable world and the potential for a changing interpretation of it. This relates to the second hypothesis, and can appear to an external observer as a small change in the external source leading to a large change in the designer's situation.
- A 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. As expectations are distributed across different layers of abstraction with links between them within and across layers, this relationship between past experiences and changes to the situation can be complex. This relates to the third hypothesis, that the organisation of knowledge arising from experience (such as within a PSS) is important for understanding why a designer makes one move (in the way that they view the design medium) and not another.

3.5.1 Testing models of interpretation

The three hypotheses about how designers move from one situation to another arise from the description of situated interpretation presented in Section 2. Implementations have been described that demonstrate these hypotheses, but the reporting of these implementations has been limited in its aim to demonstrating rather than verifying these hypotheses. The results presented here are insufficient to claim that they emulate the findings of the cognitive studies referred to in Section 1. A reason for this is that there is no established methodology for testing the situatedness of how a system interprets.

A pair of metrics can be proposed that could help facilitate verification of these hypotheses in future. The first is a metric for the *similarity* of the external source to the internal representations produced. Some relationship between the two is clearly a requirement of interpretation. The second is a metric of *reinterpretation frequency*, the frequency with which a system interprets the same source differently. It is expected that a situated system would have the capacity for frequent reinterpretation whilst having a high degree of similarity. Although this formulation is incomplete as it fails to take into account the different levels of abstraction during interpretation, however, if such metrics could be established there would be potential to conduct cognitive studies of the way that designers interpret. Then models of interpretation such as the one described and implemented here could be tested using such studies. These models of the design conversation could be termed *generate-and-interpret* in that they place the focus on producing models of design in which interpretation occurs in a designerly way.

3.5.2 Limitations of the system

A limitation of the implementations described here is that in interpretation there were occasions where neither explicit expectations in the situation, nor implicit expectations from a guided search of knowledge, could adequately construct an explanation for the data that a layer was finding. The implementations here did not include a model of situated learning, which would occur in this circumstance. These models utilised a heuristic to progressively reduce the similarity threshold until an interpretation results.

A further limitation is that the implementation 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.

The implementations described here make use of single-thread programming, however future work has the potential explore what happens when this process of interpretation occurs in parallel within each layer. This can produce a dynamic system of interpretation as described in Section 2.

4 Conclusions

This paper has conceptually linked the expectations of the designer to the process of interpretation using a computational model of design cognition. It has articulated a model for interpretation in design that accounts for the situated nature of design activity. Three hypotheses arose from the model of interpretation for situated design, and these were demonstrated through computational implementation.

The contribution is in looking towards ways that computational models of design could be made more designerly. 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 levels 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.

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 when during the process of attempting to construct an interpretation from expectations, these expectations were not met, leading to a change of expectations. This resulting change of situation, to an external observer, looks like the phenomenon of changing situations in a designer.

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.

Changes to the situation have their origin in the experiences of the designer, manifested as connections between knowledge. When the situation changes, it changes in ways that are related to the first person knowledge of the designer 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 in inspired by observed behaviour in cognitive studies. Rather than answering the original question of how designers move between situations, the work presented has raised further questions. Whilst existing cognitive studies of designers have inspired the hypotheses of this work, they only implicitly support them as the internal cognition of designers has not been reported in these studies. Future studies need to be carried out to explicitly test the way that designers move from one situation to another.

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

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

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

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

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, Stephen. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer, 21*(3), 77-88.

Carpenter, G A, & Grossberg, Stephen. (1990). ART 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures. *Neural networks, 3*(2), 129-152.

Chisholm, Roderick M. (1982). *The foundations of knowing*: University of Minnesota Press Minneapolis, MN.

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

Clancey, William. (1999). *Conceptual Coordination: How the Mind Orders Experience in Time*: LEA, Inc.

Cross, Nigel. (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, Liane, Rosch, Eleanor, & Aerts, Diederik. (2008). Toward an Ecological Theory of Concepts. *Ecological Psychology, 20*(1), 84-116. doi: 10.1080/10407410701766676

Gärdenfors, Peter. (2000). *Conceptual spaces: The geometry of thought*: Cambridge, MA, US: The MIT Press.

Gero, John S. (1998). Conceptual designing as a sequence of situated acts. *Artificial intelligence in structural engineering*, 165-177.

Gero, John S, & Fujii, H. (2000). A computational framework for concept formation in a situated design agent. *Knowledge-Based Systems, 13*(6), 361-368.

Gero, John S, & Kannengiesser, Udo. (2004). The situated function–behaviour–structure framework. *Design Studies, 25*(4), 373-391.

Gero, John S, & Smith, G. (2009). Context, situations and design agents. *Knowledge-Based Systems, 22*, 600-609.

Goel, Ashok K. (1997). Design, analogy, and creativity. *IEEE expert, 12*(3), 62-70.

Gombrich, E.H. (1966). *Norm and Form: Studies in the Art of the Renaissance*. Oxford: Phaidon.

Graf, Peter, & Schacter, Daniel 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, MD, & 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.

Hinton, Geoffrey E, Osindero, Simon, & Teh, Yee-Whye. (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, John. (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.

Kennedy, Pamela J., & Shapiro, Matthew 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, Honglak, Grosse, Roger, Ranganath, Rajesh, & Ng, Andrew 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, Marvin. (1974). A Framework for Representing Knowledge.

Montare, Alberto. (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, VB. (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, Allen. (1994). *Unified theories of cognition* (Vol. 187): Harvard University Press.

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

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.

Piaget, J. (1954). *The Construction of Reality in the Child*: Basic Books.

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.

Pylyshyn, Z W. (1984). *Computation and Cognition*: Harvard University Press.

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

Rosch, Eleanor. (1978). Principles of Categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale: Lawrence Erlbaum Associates.

Schacter, DL. (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, DW, Hatfield, D, Svarovsky, GN, 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, Masaki, Gero, John, & Purcell, Terry. (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, Masaki, & Tversky, B. (1997). What do architects and students perceive in their design sketches? *Design Studies, 18*(4), 385-403.

Thoreau, Henry David. (1851/1993). *A Year in Thoreau's Journal: 1851*. New York: Penguin Books.

Tscherepanow, Marko. (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.

Yaner, Patrick W, & Goel, Ashok K. (2008). Analogical recognition of shape and structure in design drawings. *AI EDAM, 22*(2), 117-128.