# COGNITIVE BIASES AND PRECEDENT KNOWLEDGE IN HUMAN AND COMPUTER-AIDED DESIGN THINKING

JOO-HWA BAY
*National University of Singapore, Assistant Professor, Dr.*
*Email address: akibayp@nus.edu.sg*

**Abstract.** Cognitive *biases* (illusions) and potential errors can occur when using precedent knowledge for analogical, pre-parametric and qualitative design thinking. This paper refers largely to part of a completed research (Bay 2001) on how heuristic *biases*, discussed by Tversky and Kahneman (1982) in cognitive psychology, can affect judgement and learning of *facts* from precedents in architectural design, made explicit using a kernel of conceptual system (Tzonis et. al., 1978) and a framework of architectural representation (Tzonis 1992). These are used here to consider how such illusions and errors may be transferred to computer aided design thinking.

## 1. Background

In the use of precedents as analogy for creative problem solving in design, implicit assumptions are often made about the success of precedents and the possible mapping over of the structure of *morphology*, *operation*, and *performance*[1] to new design problems in a new contexts. The selection of precedent knowledge for used in computer retrieval and reasoning systems are often based on human judgments and beliefs that these precedents are successful. Most judgements and descriptive statements of the success of historical and contemporary architectural examples are based on human heuristics rather than normative assessments because of the limitations of computational resources and time in practice.[2] As human heuristics can be

[1] Refer to Tzonis (1992) for more about representing architectural knowledge for analogical thinking creative thinking. The author here uses the concepts and relationships of *morphology*, *operation* and *performance* for discussion throughout this paper.
[2] Please refer to Bay (2001, 49-72) for more discussion on use of precedents, beliefs, judgement and decision-making in practice in the case of tropical architecture.

affected by associated cognitive *biases* (illusions)[3] that lead to systematic errors in design thinking (Bay 2001), knowledge contents in computer knowledge systems can also be saddled with in-built potentials for errors in application.[4]

For instance, in the use of traditional houses as precedent knowledge, assumptions are made for their success and applicability across context of past environment to present urban settings (Bay 2001, 46). Also in another example, in Fang (1993) the choice of using the prototype Beijing courtyard house as the precedent knowledge for a computer precedent analysis to facilitate design reasoning is based on the assumption and belief that the new courtyard house operates and performs well, and that the knowledge can be used as facts for computer learning. These beliefs are made with human heuristics with little or no recourse to normative computational methods of assessments, and may be subjected to cognitive biases (illusions).

In computer precedent retrieval and reasoning systems, the human approach to using precedents for problem solving is emulated, and with it also the heuristic assumptions and beliefs in the precedent knowledge and application. As a result, the knowledge content and usage may be affected by heuristic biases that lead to errors, transferred from the human cognitive systems to the computer systems.

## 2. Human learning of dysfunctional facts

The actual internal design thinking process in the mind of the designer is not totally obvious, but can to a certain degree be described and understood with a model based on the external thinking process, observed through the descriptive and prescriptive (normative) statements made by the designer. A Kernel of Conceptual System represents a minimal necessary cognitive structure of argumentation based on the theory of action, and evaluation of action (Tzonis et al., 1978). This is used with a framework for representing architectural knowledge in analogical creative thinking (Tzonis 1992) to model an instance of design thinking in tropical architecture in Figure 1.

[3] Cognitive '*bias'* used here is in the sense referred to by Tversky and Kahneman (1982) and Osherson (1995) as 'illusion' related to certain heuristic used in thinking that leads to errors in judgement and decision making. This may not be confused with 'bias' (preference) in the sense that 'he has a bias for red', meaning 'he tends to prefer red emotionally or ideologically as a colour or style'. This does not mean that his 'tendency' or 'preference' cannot cause a heuristic *bias*; i.e. if this inclination influences a heuristic in use, a resulting heuristic *bias* can be linked to it.

[4] *Biases* are not the only reasons for incidences of design judgement errors. Errors can be attributed to many other factors, including poor education, low fees that limit search for solutions, outdated building codes and regulations, and insufficient details about precedents in reading materials. In a sense they are interrelated and there are overlaps in their effects on design thinking. This paper focuses on the effects of *biases.*

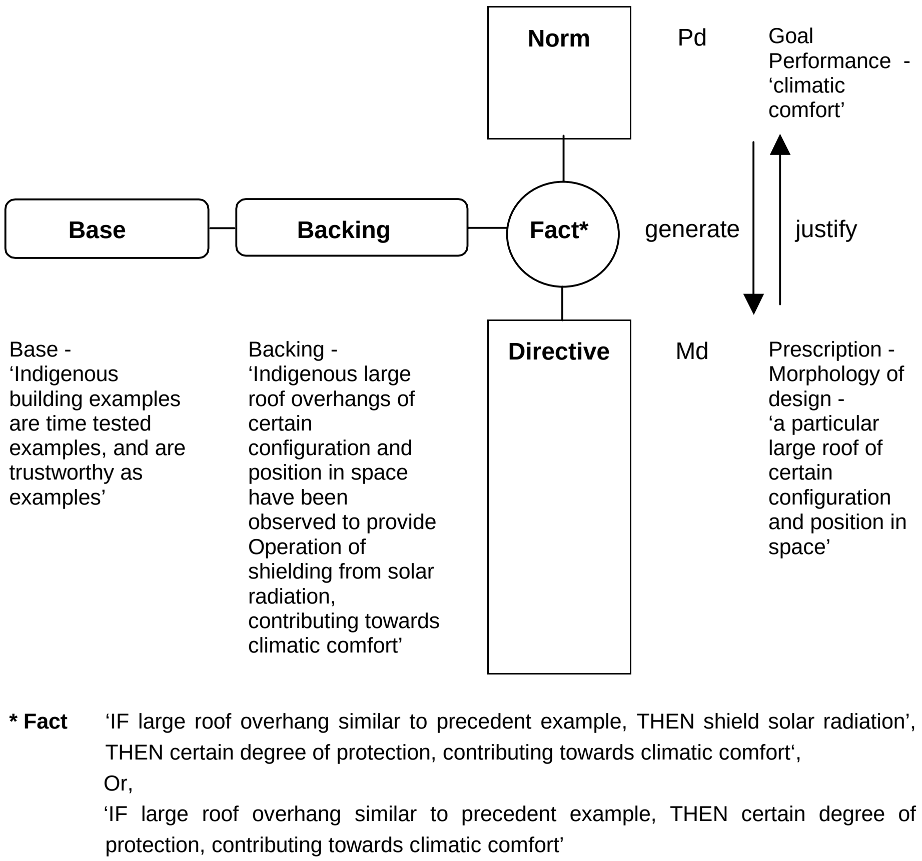


*Figure 1.* Kernel of Conceptual System; with an example for a roof morphology directive for a 'climatic comfort' norm, justified by a fact, learned from a backing and base

In the instance (Figure 1), the learning of the *fact* 'IF large roof overhang similar to precedent example, THEN…. certain *performance*', heuristics are used for judgement. One of these heuristics employs the mechanism of similarity (*representativeness* or match) of problem structure for reasoning. The association with success of the *backing* example is another heuristic used dependent on the *availability* of authority and trustworthiness of the example; i.e. the *base*. Systematically, a *norm* (goal) for a certain degree of goal *performance* (Pd) relating to this *fact* should generate a corresponding prescriptive *morpholog*y (Md) of a certain physical attribute, which when built should perform as expected. However, if there are cognitive *biases* due to *representativeness* and *availability*, they can cause unwarranted confidence of judgement in the *fact*. Since the decision for a certain prescribed *morphology* intended for implementation depends on the faulty *fact* statement, then the actual *performance* may not be as desired when built.

For instance, the roof may look like it will work, but actually do not work well, but the *representativeness bias* creates the overconfidence in the judge to assess that it works well corresponding to the degree it is *representative* of success. Another instance, the roof does not work well, but because the example is well publicised, the *availability bias* creates overconfidence in the judge to think that it works well corresponding to the degree of fame or salience making it easier to imagine success. If any of these happens, the *fact* has an in-built error due to *bias*.

According to Rescher (1966) for Heterogeneous Command Inference, where command means *norms* here, “A command inference that infers a command conclusion from premises containing a mixture of commands and assertoric statements can be ‘valid’ only if the command conclusion must be terminated whenever (i.e., in any possible world in which) all the command premises are terminated and all of the assertoric premises are true”. ‘Assertoric premises’ here refers to the descriptive *fact* statement, and it must be true for the *directive* to be ‘valid’. If the desired *performance* (Pd) generates a corresponding desired *morphology* (Md), justified by a dysfunctional *fact* (belief or descriptive statement), then it is not ‘valid’.

The cognitive *biases* related to *representativeness and availability*, are termed by Kahneman and Tversky (1982) as the ‘*illusion of validity*’, and *biases due to imaginability* respectively, and can affect professionals in the financial, legal and clinical context. Cognitive experimentation (Bay 2001, 155-170) shows that subjects given photographs of precedent architectural projects can be overconfident in judging the *performance* of these examples because they looked like they will perform well but actually do not in reality. Also separately because they were projects by famous architects making it easier to imagine that they perform more than they actually can. Case study of design judgements by architects and writers (op. cit.) also shows that professional, successful and famous architects and writers can made mistakes related to *representativeness* and *availability biases*.

Many architectural publications of buildings carry descriptive statements of *performances* of buildings that are based on the writer’s heuristic judgement. Judgement is usually made without normative quantitative calculation. The spatial quality assessments and predictions are done with the mind and senses[5], usually by comparing graphically and visually with past experience and precedents. There are possibilities of errors due to cognitive *biases* embedded in these selections of so-called ‘good

[5] This includes surveying a building on site, commonly known as ‘visual inspection’ of conditions, as opposed to normative quantitative measurements with instruments and calculations.

precedents’[6], which can be problematic. This type of precedent knowledge in literature, with embedded errors, not only cause errors in human design thinking, but they can transfer the errors due to *biases* to other literature and computer knowledge systems that use them as knowledge base, and perpetuate the problem.

## 3. Computer-aided knowledge retrieval and machine learning of facts

The problem above is particularly applicable to the case of creative-associative knowledge retrieval systems, like the ‘OASIS’ (Tan et. al 2000), or other simpler visual collection information retrieval systems, where the knowledge input are judged in pre-parametric (Ulrich 1988) and qualitative (Kuipers 1994) modes, and usually with limited descriptive or belief statements about the *performances* and usually lack of statements about the *operations* of various *morphologies* in the visual information.

In the Architectural Precedent Analysis (ARPRAN) approach, Fang (1993) used the precedent knowledge based on literature that reported or claimed the success of *performance* and *operation* of the *morphology* of the new Beijing courtyard house, which were designed based on heuristic usage of precedent knowledge of older courtyard houses and the city. Fang assumed those *facts* were accurate, but the likelihood of errors due to *biases* could be there, and machine learning from this problem solving experience might be problematic.

Similar systems to precedent analysis are systems that employ example-based learning (Fang 1993, 116), which requires both similarity-based and explanation-based methods to learn from examples. A system developed by Kedar-Cabelli (1988) will be used here to discuss how cognitive *biases* can affect computer example-based learning within the structure of the system.

In the *Computational model of purpose-directed analogy* (Op. cit., 96-102), Kedar-Cabelli gave a case study of generalisation using the interaction of goal concept, operations and physical properties to explain a ‘HOT-MUG’ with a base ceramic mug and to justify a target styrofoam cup as a ‘HOT-MUG’, and in turn use this knowledge (*fact*) learned as a generalisation of a concept definition of a ‘HOT-MUG’.[7] In the case, the plan of action leading to the main purpose ‘enable INGEST hot-liquid’, includes enabling operations, such as IF PUTIN, KEEP, GRASP, PICKUP, INGEST, THEN ‘enable INGEST hot-liquid’. These enabling operations are in turn connected to preconditions, IF ‘can (be-grasped-by agent, hot-liquid),

[6] ‘Good precedents’ here includes examples and related rules, and even typology in a general sense that is belief to be adequate and suitable to guide design in new problems.

[7] In a sense, there is similarity to argumentation in the Kernel of Conceptual System where the backing example provides concept structure and learning of a *fact*, and generalisation.

IF structural features, IF 'insulate-heat grasping-area', and attribute, IF 'material (ceramic) has-part (handle)'.

Relating to arguments above in this paper, these are interrelated descriptive or belief statements, which are input or supplied by the system developer (human). The IF-THEN statements relate these elements together as if belief statements or *facts.* Could these be subjected to possible heuristic *biases*? In the generalisation there is a statement derived from the target and the explanation in the base, which goes like this: 'IF styrofoam, insulate-heat, grasping-area, THEN can be grasped by agent, THEN enable INGEST hot-liquid'. It is believed that the conical shape of the styrofoam material has a 'grasping-area' therefore it is graspable. However, if the styrofoam is too thin, but the conical shape looked like graspable, then the *fact* "IF styrofoam, conical, THEN graspable-area, THEN etc' could suffer a *bias* of *representativeness* transferred from the human supplier of related descriptions. If such an error happens, then it becomes embedded in the knowledge system.

Can the human supplier of related descriptions for the computer be aware that he is suffering a *bias* without external prompting? If he knows that he may suffer from *biases*, can that knowledge itself help eliminate the problem? According to Arkes (1986, 587), "One technique that has proven to be absolutely worthless is telling people what a particular *bias* is and then telling them not to be influenced by it". This is because the subjects do not actually know how to overcome these *biaes* even if they know they may exist and affect them. Bay (2001, 168) also showed that the warning about the existence of both *representativeness* and *availability biases* was no help in *debiasing* in the case of design thinking in tropical architecture.

## 4. Therapeutic mechanism and possible applications

For *debiasing* in human design thinking, Bay (Op. cit.) tested and showed that the introduction of *rebuttal* to increase *availability* of thoughts of opposite outcomes can refocus the attention of the mind making judgement, and can improve accuracy of judgement subjected to *illusion of validity* and *biases due to imaginability*.

Possible application could be to increase the *availability* of design knowledge of buildings where they do not work; to put 'how and why' they do not work alongside information that they work, and with the varying degrees of success assessed with normative methods. This can possibly serve as the *rebuttal-debiasing* mechanism for the human precedent user to reduce errors because of embedded *biases* in *facts* (or concept definition for purpose-directed analogy). This could also help in improving the selection of more accurate precedents for computer knowledge retrieval systems and for machine learning. Fang (1993, 108), referring to Carbonell (1983), noted that similarity-based learning can be classified as learning from only positive examples and learning from both positive and negative examples, where learning from positive examples only do not provide information for preventing over-generalisation of the inferred concept. He asserted that over-generalisation for learning from only positive examples might be avoided by considering only the minimal necessary generalisation, or by relying upon a prior domain knowledge to constrain the concept to be inferred. Carbonell (1995, 591) also noted that analytic systems utilise "past problem solving experience (the exemplars) to guide which deductive chains to perform when solving new problems", thus increasing efficiency without sacrificing accuracy and generality. However, as argued above, the prior knowledge and the exemplars itself may be flawed with transferred and embedded *biases* and potential errors. As for providing both positive and negative examples for inductive computer learning to prevent over-generalisation, this can be considered for improving accuracy of machine learning, provided some normative checks are made to ensure positive examples are truly positive and negative examples are truly negative, and not just look like they are negative but actually works. Several analytic systems that allow learning from successes and failures (Op. cit. 589-591) could similarly be considered for improving machine learning accuracy and limit the problem of possible embedded errors due to *biases*.

## 5. Conclusion

Human beings use heuristics to quickly solve complex problems. However, cognitive *biases,* associated with heuristics, can affect human design thinking in practice, with possible systematic errors. Errors due to *biases* can be transferred through precedent knowledge and *facts* in computer-aid retrieval and machine learning systems, and become embedded errors.

Carbonell (Op. cit. 589, 594) noted new systems that explore the possibility of inaccurately labeled and unlabeled instances for the inductive machine learning, and the possibility of combining inductive and analytic systems. To what extent can these systems prevent or reduce possible transferred errors due to *biases*? This is an interesting area to explore, as computers continue to be more human-like in heuristics.

## Acknowledgement

This paper refers largely to the author's completed Ph.D. study at and enabled by the Design Knowledge System Research Centre at Delft University of Technology, with the generous grant of study leave with financial help from the National University of Singapore. Much gratitude goes to Professor Alexander Tzonis as the Ph.D. supervisor for his insightful guidance, and the colleagues at DKS Research Centre for their stimulating discussions.

## References

Arkes, Hal R.: 1986. Impediments to accurate clinical judgment and possible ways to minimize their impact, in Arkes, Hal R., and K. Hammond, (eds), Judgment and decision making, Cambridge University Press, USA, pp. 582-592.

Bay, Joo-Hwa: 2001. Cognitive biases in design: The case of tropical architecture. The Design Knowledge System Research Centre, Faculteit Bouwkunde, Technische Universiteit Delft, The Netherlands.

Carbonell, J. G.: 1983. Derivational analogy and its role in problem solving, in, Proceedings AAAI-83, Washington, D.C., pp. 45-48.

Carbonell, J. G.: 1995. Paradigms for machine learning, in Luger, G. F. (ed.), Computation & Intelligence: Collected readings, AAAI Press, The MIT Press, pp. 587-595.

Fang, Nan: 1993. A knowledge-based computational approach to architectural precedent analysis. Publikatieburo Boukunde, Technische Universiteit Delft, The Netherlands.

Kedar-Cabelli, S.: 1988. Toward a computational model of purpose-directed analogy, in Armand Prieditis (ed.), Analogica, Morgan Kaufmann Publishers, Inc, California; Pitman, London, pp. 89-107.

Kuipers, Benjamin: 1994. Qualitative reasoning: Modelling and simulation with incomplete knowledge. MIT Press, Cambridge, Mass.

Osherson, Daniel. N.: 1995. Judgment, in E. E. Smith, and Daniel N. Osherson (eds), Thinking, Invitation to cognitive science, Vol. 3, MIT Press, Cambridge, MA, pp. 55-87.

Rescher, Nicholas: 1966. The logic of commands. Routledge and Kegan Paul, London.

Tan, M., B. K. Tan, and J. Ngatemin: 2000. By rhyme or reason: Rapid design thinking by digital cross referencing, in B. K. Tan, M. Tan, and Y. C. Wong (eds), Proceedings of the fifth conference on computer aided architectural design research in Asia: CAADRIA 2000, published by CASA, National University of Singapore, pp. 399-410.

Tversky, A., and D. Kahneman: 1982. Judgment under uncertainty: Heuristics and biases, in D. Kahneman, P. Slovic, and A. Tversky (eds), Judgment under uncertainty: heuristics and biases, Cambridge University Press, Cambridge, UK, pp. 3-22.

Tzonis, Alexander: 1992. Huts, ships, and bottlenecks: Design by analogy for architects and/or machines, in Nigel Cross, Kees Dorst, and Norbert Roosenburg (eds), Research in design thinking, TUDelft, Delft, pp. 130-165.

Tzonis, A., R. Berwick and M. Freeman: 1978. Discourse analysis and the logic of design, in Alexander C. Tzonis (ed.), Conceptual systems in design: Two papers, Harvard Graduate School of Design Publication Series in Architecture, A: 7817.

Ulrich, Karl T: 1988. Computation and pre-parametric design, Technical report of MIT Artificial Intelligence Laboratory. AL-TR 1043.