Making Rebuttals Available Digitally for Minimising Biases in Mental Judgements

Joo-Hwa BAY
Department of Architecture, National University of Singapore, Singapore

Keywords: analogy, bias, design thinking, environmental design, heuristics

Abstract: The problem of heuristic biases (illusions) discussed by Tversky and Kahneman (1982) that can lead to errors in judgement by human designers, when they use precedent knowledge presented graphically (Bay 2001). A Cognitive framework of belief, goal, and decision, and a framework of representation of architectural knowledge by Tzonis are used to map out the problem of heuristic biases in the human mind. These are used to discuss what aspects of knowledge can be presented explicitly and digitally to users to make rebuttal more available for human thinking at the cognitive level. The discussion is applicable to both inductive and analytic digital knowledge systems that use precedent knowledge. This discussion is targeted directly at means of addressing *bias* in the *human* mind using digital means. The problem of human bias in machine learning and generalisation are discussed in a different paper, and the problems of international or non-intentional machine bias are not part of discussion in this paper.

1 INTRODUCTION

This paper focuses on the problem of heuristic *biases* (biases as illusions, Tversky and Kahneman 1982; Osherson 1995)[1] that can lead to unwarranted confidence and errors in judgement by *human* designers, when they use architectural precedent knowledge for tropical climatic designs (Bay 2001). This is different from "intentional bias" and "explicitly biased" generalisation for improving machine learning (Rosenbloom et. al 1993; Gordon et. al 1995). This paper refers to *biases* as illusions that can be embedded in precedent knowledge (rules and principles as well as example or case) in publications and research data, and can also be transferred into various forms of digital knowledge systems, as they are overconfidently assumed accurate at point of input by human judges (Bay 2002) [2]. Feedbacks

[1] This may not be confused with 'bias' (preference, slant) 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 'inclination' cannot cause a heuristic *bias*; i.e. if this inclination influences a heuristic in use, a resulting heuristic *bias* can be linked to it.

[2] For more discussion on this, please refer to Bay (2002) which discussed possible *biases* affecting

solicited from human experts, like “similarity teacher” (Stahl 2001), can also be subjected to *biases* and embedded errors.

This paper discusses what aspects in knowledge system can be made explicit and available to designers on the *human* cognitive level for minimising un-intentional and un-desirable *human mental biases*. It does not discuss the mechanism of how machine learn more accurately, including: algorithms to facilitate improvement of accuracy of machine retrieval by learning from failures (Carbonell 1990, Kolodner 1993, 234, Cox 1997), integration of analogical search control and explanation-based learning (VanLehn et. al. 1993), using statistical probability of correctness, learning from multiple sources of inaccurate data (Baliga et. al. 1992, Carbonell 1990), using truth maintenance system with convenience of changes (Doyle 1995), and learning from positive and negative examples (Börner 1998, 207, Carbonell 1990).

It is about boosting ‘truth’ accuracy in the *minds of human* designers (Figure 1) when he is using the digital precedent knowledge for various applications in new design problems (the process of doing this may or may not boost machine ‘truth’ accuracy). With representations of the nature of how this human mental *biases* comes about and the possible *rebuttal* mechanism for *debiasing*, this paper will suggest what aspects in digital systems can be make explicit to the *human* mind, in order to increase availability of *rebuttal* to counteract overconfidence in the designer.

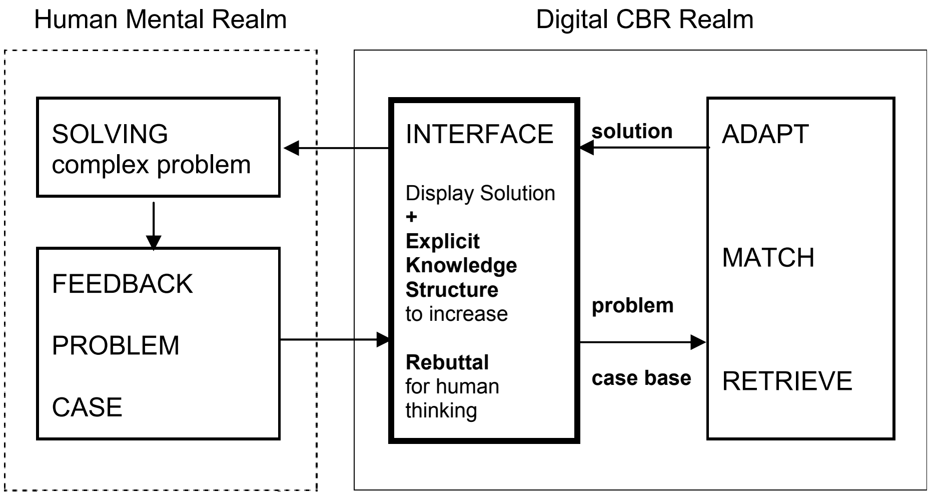


Figure 1 Display explicit knowledge structure to increase rebuttal to human thinking for problem solving, input of case base, problem and feedback.

---

retrieval, justification, and explanation systems by Milton et al. 2000, Kedar-Cabelli 1988, Fang 1993.

2 ASSUMPTIONS

Generalised design knowledge are usually suited for clearly defined problem space, and seldom (at least for now) used to automate processes of innovation and creativity in complex domains, such as multi-criteria architectural design, as observed by Börner (1998). As such CBR systems are mainly used as 'design assistant' (Op. cit.), with much useful knowledge learnt from cases, and retrieved in the best possible way to help architectural designers, rather than automate the design process (Figure 1 shows CBR system as an assistant).

Such systems are usually used to provide the best-case adaptation, justification, and generalisation for the users to make the final creative analogical mapping and transformation. In reality, robust cases are rare, and most examples are incomplete and may be inaccurate in representation. Machine can be wrong due to embedded *biases* and errors. These flaws and gaps in knowledge structure are usually not explicit.

For creative design, human designers use many different precedents, and not necessarily those that the computer knowledge system proposed as most accurate. Therefore, the problem with *biases* can affect design. Also, even with the use of accurate machine knowledge, and the mapping of knowledge over to new problem *context*, transfer and transformation may be affected by *human* mental heuristic *biases.*

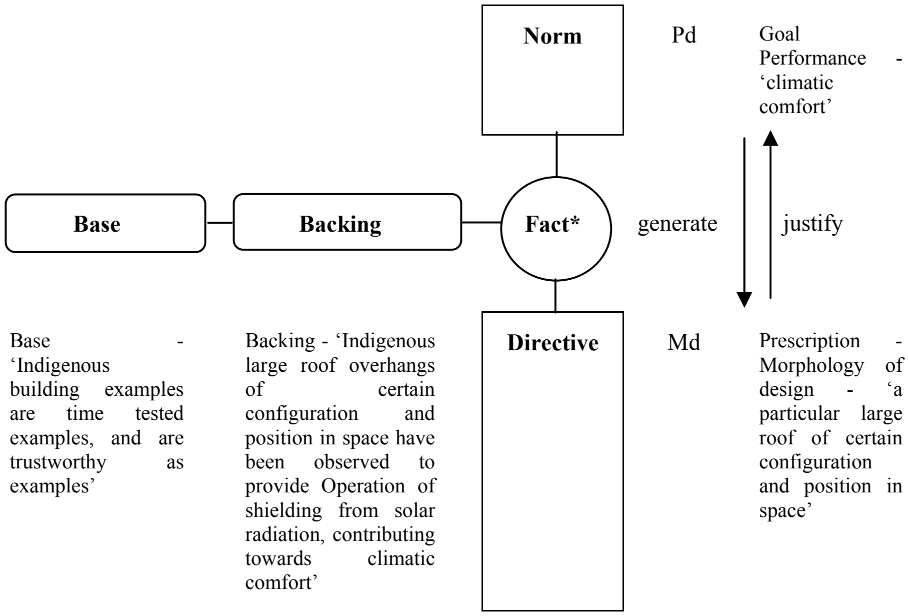
3 COGNITIVE BIASES AND HUMAN DYSFUNCTIONAL JUDGEMENT

This section summarises the dysfunction judgement of architectural designers handling knowledge in tropical environmental design owing to heuristic *biases* discussed in Bay (2001, 2002).

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. The Kernel of Conceptual System by Tzonis et al. (1978) can be used to represent a minimal necessary cognitive structure of argumentation based on the theory of action, and evaluation of action. This is used together with a framework for representing architectural knowledge in analogical creative thinking (Tzonis 1992), with representational concepts of *performance, operation, morphology,* and *context* to model an instance of design thinking in tropical architecture in (Figure 2).

In the instance (Figure 2), heuristic judgements are used in learning of the *fact* 'IF large roof overhang similar to precedent example, THEN…. certain *performance*'. One of these heuristics employs the mechanism of similarity (*representativeness* or match, Kahneman and Tversky 1982) of problem structure for reasoning. The association with success of the *backing* example is another heuristic used dependent on the *availability* (Kahneman and Tversky 1982) 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*.



* Fact 'IF large roof overhang similar to precedent example, THEN shield solar radiation', THEN certain degree of protection, contributing towards climatic comfort',
Or,
'IF large roof overhang similar to precedent example, THEN certain degree of protection, contributing towards climatic comfort'

Figure 2 Kernel of conceptual system; with an example for a roof morphology, with directive for a 'climatic comfort' norm.

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, they made it easier to imagine that they perform more than they actually can. Case study of design judgements by architects and writers (op. cit. 110-116) also shows that professional, successful and famous architects and writers can made mistakes related to *representativeness* and *availability biases*.

Errors can also arise in wrongly believing that the *norm* (goal) matches the *fact* because of similarity (*representativeness*) and is an appropriate precedent for analogical transfer and transformation of dimension to another context.

4 DEBIASING

For *debiasing* in human design thinking for tropical architecture, 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, and can improve accuracy of judgement that are subjected to *illusion of validity* and *biases due to imaginability*.

In Bay’s experiment (Op. cit., 161-162), a sectional view of a project is shown, highlighting the negative *operation* and *performance* of sun-shading, making more available the possibility of failure as the *rebuttal* and therapeutic input. Increasing the *availability* of thoughts on ‘how and why’ they may not and do not work, improve human judgement.

How can this be applied in the case of precedent knowledge in digital system, and what can the system make explicit rebuttals to increase chances of *human debiasing*?

5 MAKING REBUTTALS AVAILABLE

5.1 Negative Examples

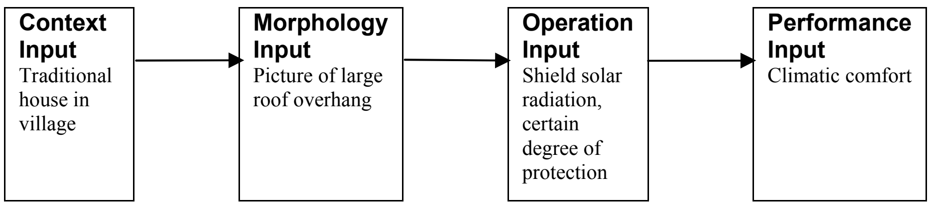
Human learners also can improve accuracy in learning of facts by learning from both positive and negative examples. Similarly, machine learning of positive and negative examples can improve the accuracy of learning facts for the machine. For example, Börner (1998) and Carbonell (1990) discussed about improving machine learning for accuracy by learning from good positive and negative cases. In a sense increasing knowledge of both positive and negative examples can reduce the probability of over generalization for both machine and human.

However, the similarity breaks down because the *human* mind is more complex than the machine, and is subjected to *psychological* illusions like cognitive *biases*. The highlighting of opposite outcomes as *rebuttals* is necessary to reduce certain cognitive *biases* pertaining to the human user. Besides showing negative examples, this paper proposes that digital knowledge system can make explicit, and in reverse, the structure of causal links and elements of the knowledge, to strengthen the rebuttal for *human debiasing*.

5.2 Explicit Structure of Knowledge

Cox (1997) discussed about making explicit representation of reasoning failure in machine, which in a way helps to show the machine's mistakes, but it does not show the possible mistakes by the user at input or using the knowledge for mental thinking.

Knowledge from precedents is not necessarily complete in reality, digital or not. Usually the machine learning process, learned structure of causal links and elements of the knowledge in each case is transparent to the user, but is available only to the machine for computation of retrieval priorities. Showing the incompleteness of the learned structure can alert users of missing elements or links, and therefore increase availability of human thoughts on possible negative conclusions.

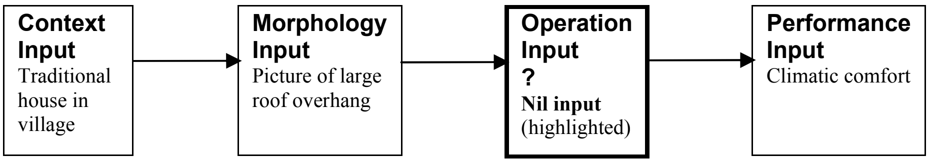


Knowledge learnt: 'IF large roof overhang similar to precedent example, THEN shield solar radiation', THEN certain degree of protection, contributing towards climatic comfort' for traditional house in village.

Figure 3 Context-Morphology-Operation-Performance Knowledge Structure

In the example above in Figure 2, the *context-morphology-operation-performance* structure of the precedent could take two forms. It can be a full structure like in Figure 3, or one that is incomplete in Figure 4.

Showing incomplete structure and missing element, for instance, missing *operation* (Figure 4), alerts the user that conclusion is made without information and analysis of the *operation*, and the conclusion can be wrong. This type of explicit displays to the user makes available the cognitive attention on possible failure of generalisation, and forms the *rebuttal* for *debiasing*.



Knowledge learnt: 'IF large roof overhang similar to precedent example, THEN climatic comfort' for traditional house in village.

Figure 4 Incomplete Knowledge Structure Made Explicit

5.3 Sectional Views for Climatic Design

In Bay's (Bay 2001) experiment with the knowledge domain of tropical climatic design, sectional views are used to increase availability of negative outcome to the judges. For this particular category of design knowledge, it is recommended that sectional views be part of the necessary input and visual display to increase *debiasing* potentials for this particular knowledge domain.

In the case where no sectional view is available, this can also be made explicit and highlighted to the users that they are not viewing a sectional view and do not have the advantage of visually assessing the positive and negative *operation* of shading for instance.

5.4 Distance in Context Compatibility

In analogical thinking, knowledge learned from one *context* is mapped over to another. Sometimes the context of the precedent is dramatically different from the new *context* of the design problem. For instance, designing a house in the dense urban city base on knowledge learned from a traditional house in the village. Mere transfer of the structure and straightforward dimensional transformation for the new design from the old example may be an overconfident judgement, based on the similarity of house design and *morphology,* especially if the similar roof form is used.

If the machine can make explicit the distance in the context compatibility besides the structural similarity, this will also help alert the possibility of overconfidence, and *debias*. It can also be helpful to make available projections of the likelihood of success, by way of the comparison of distance/similarity and structure[1] of the knowledge to new context.

5.5 User Feedback

Many systems include user feedback as a rich resource of adding accuracy to the knowledge base (Stahl 2001). Human user's experience, of both correct and incorrect knowledge inference, can be good feedback to the system to improve accuracy. Also, as mentioned earlier, the human mind judging the knowledge at input point can be subjected to cognitive *biases* and thus transfer embedded errors in the knowledge system. The explicit system proposed above to increase rebuttal can also be used at point of input to alert the human user in order to avoid such errors.

6 DISCUSSION AND CONCLUSION

The creative architects looks at precedent cases for multi-criteria analogical thinking that has many conflicting priorities and may not necessary use the best example of climatic design for inspiration for his complex design. It is also more accurate to use parametric computations and simulations (requiring expert advisors to the architect) for the new design, but this is usually not possible for the creative designer, who because of the complexity of projects, limited time and budgets, necessitates him to use pre-parametric (Ulrich 1988) heuristic means (shortcuts and rules-of-thumb) of design judgement and decisions.

CBR systems with algorithms that retrieve the best examples may help solve the problem of not getting the most accurate example. However, it does not solve the problem of *human* mental *biases* in the mapping over for new problem, with overconfidence in the aptness of mapping and transformation.

Bay (2001) has shown that *rebuttals* can actually *debias* and improve heuristic judgement. *Rebuttals* can be made more available by showing explicit structure of knowledge and the incompleteness of causal elements. Also in this specific knowledge for tropical design, by showing sectional views, and if not available, to highlight the fact that they are not available for visual checking of *operation* for climatic controls. Distance and statistical weightings, and projected effects can also be made explicit and available. These can also be made available at point of input and feedback by users to system. However, there must be a balance of key information, and *rebuttal* cues for the user.

As mentioned, the basic aim of CBR is to be an architect's assistant (Börner 1998), matching and finding best-fit structure and features for the user to make creative

[1] For example of comparison of distance, see 'projecting effects and Battle plans' by Goodman, discussed by Kolodner, Janet L., and David B. Leake (1996).

judgements and decisions. If it is too complicated, then the heuristic creative designer will be too discouraged to use the machine to help him.

REFERENCES

Baliga, G., S. Jain and A. Sharma. 1992. Learning from Multiple Sources of Inaccurate Data. *Analogical and Inductive Inference, Proceedings of International Workshop AII ' 92, Dagastubhl Castle, Germany, October 1992,* ed. K.P. Jantke. LNAI 642: 108-128. Germany: Springer-Verlag.

Bay, J.H. 2001. *Cognitive Biases in Design: The Case of Tropical Architecture.*, The Netherlands: The Design Knowledge System Research Centre, Faculteit Bouwkunde, Technische Universiteit Delft.

Bay, J.H. 2002. Cognitive Biases and Precedent Knowledge in Human and Computer-aided Design Thinking.

Proceedings of CAADRIA 2002, eds. Eshaq, Khong, Neo and Ahmad, 213-220. Selangor, Malaysia: Prentice Hall. *Proceedings of CAADRIA 2002*, eds. A. Eshaq, W.K. Chee, K.T.K. Neo, M. Neo, S.N.S. Ahamad, 213-220. Selangor, Malaysia: Prentice Hall.

Börner, K. 1998. CBR for Design. *Case-based Reasoning Technology: From Foundations to Applications,* eds. M. Lenz, B. Bartsch-Sporl, H. Burkhard, and S. Wess. LNAI 1400: 201-233. Germany: Springer.

Carbonell, J.G. 1990. Introduction: Paradigms for Machine Learning. *Machine learning: Paradigms and methods,* ed. J. G. Carbonell, 1-10. Cambridge, Mass: The MIT Press.

Cox, M.T. July 25-27, 1997. An Explicit Representation of Reasoning Failures. *Case-based Reasoning Research and Development, Proceedings of 2nd International Conference on Case-Based Reasoning. ICCBR-97. Providence, RI, USA*, eds. David B. Leake and Enric Plaza. LNAI 1266: 211-222. New York: Springer.

Doyle, J. 1995. A Truth Maintenance System. *Computation & Intelligence: Collected Readings,* eds. George F. Luger, 529-554. AAAI Press/ The MIT Press.

Gordon, D. and D. Perlis. 1995. Explicitly Biased Generalization. *Goal-driven Learning,* eds. Ashwin Ram and David B. Leake, 321-354. Cambridge, Mass: The MIT Press.

Kolodner, J.L. 1993. Case-based Learning. *Reprint from A Special Issue of Machine Learning,* vol. 10, No. 3. Boston/Dordrecht/London: Kluwer Academic Publishers.

Kolodner, J.L., and D.B. Leake. 1996. A Tutorial Introduction to Case-based Reasoning. *Case-based Reasoning: Experiences, Lessons, & Future Directions,* ed. D. B. Leake, 31-66. AAAI Press/The MIT Press

Rescher, N. 1966. *The Logic of Commands.* London: Routledge and Kegan Paul.

Rosenbloom, P.S., S. Lee and A. Unruh. 1993. Bias in Planning and Explanation-cased Learning. *Machine Learning Methods for Planning,* ed. Steven Minton, 197-232. San Mateo, California: Morgan Kaufmann Publishers.

Stahl, A. 2001. Learning Features Weights from Case Order Feedback. *Case-based Reasoning Research and Development, Proceedings of 4th International Conference on Case-Based Reasoning, ICCBR 2001,* eds. David W. Aha, and Ian Watson. LNAI 2080: 502-516. New York: Springer.

Tversky, A. and D. Kahneman. 1982. Judgment under Uncertainty: Heuristics and Biases. *Judgment under Uncertainty: Heuristics and Biases*, eds. D. Kahneman, P. Slovic and A. Tversky, 3-22. Cambridge, UK: Cambridge University Press.

Tzonis, A., R. Berwick and M. Freeman. 1978. Discourse Analysis and the Logic of Design. *Conceptual Systems in Design: Two Papers,* ed. Alexander C. Tzonis. Harvard Graduate School of Design Publication Series in Architecture, A: 7817.

Tzonis, A. 1992. Huts, Ships, and Bottlenecks: Designed by Analogy for Architects and/or Machines. *Research in Design Thinking,* eds. Nigel Cross, Kees Dorst, and Norbert Roosenburg, 130-165. Delft: TUDelft, Delft.

Ulrich, K.T. 1988. Computation and Pre-parametric Design. *Technical Report of MIT Artificial Intelligence Laboratory.* AL-TR 1043.

VanLehn, K. and R.M. Jones. 1993. Integration of Analogical Search Control and Explanation-based Learning of Correctness. *Machine Learning Methods for Planning,* ed. Steven Minton, 273-316. San Mateo, California: Morgan Kaufmann Publishers.