Relevancy in Problem Solving: A Computational Framework

Johan Kwisthout[1]

Abstract

When computer scientists discuss the computational complexity of, for example, finding the shortest path from building A to building B in some town or city, their starting point typically is a formal description of the problem at hand, e.g., a graph with weights on every edge where buildings correspond to vertices, routes between buildings to edges, and route-distances to edge-weights. Given such a formal description, either tractability or intractability of the problem is established, by proving that the problem either enjoys a polynomial time algorithm or is NP-hard. However, this problem description is in fact an abstraction of the actual problem of being in A and desiring to go to B: it focuses on the relevant aspects of the problem (e.g., distances between landmarks and crossings) and leaves out a lot of irrelevant details.

This abstraction step is often overlooked, but may well contribute to the overall complexity of solving the problem at hand. For example, it appears that "going from A to B" is rather easy to abstract: it is fairly clear that the distance between A and the next crossing is relevant, and that the color of the roof of B is typically not. However, when the problem to be solved is "make X love me", where the current state is (assumed to be) "X doesn't love me", it is hard to agree on all the relevant aspects of this problem.

In this paper a computational framework is presented in order to formally investigate the notion of relevance in finding a suitable problem representation. It is shown that it is in itself intractable in general to find a minimal relevant subset of all problem dimensions that might or might not be relevant to the problem. Starting from a computational complexity stance, this paper aims to contribute a computational framework of 'relevancy' in problem solving, in order to be able to separate 'easy to abstract' from 'hard to abstract' problems. This framework is then used to discuss results in the literature on representation, (insight) problem solving and individual differences in the abstraction task, e.g., when experts in a particular domain are compared with novice problem solvers.

Keywords

relevancy, abstraction, computational complexity, formal modeling, problem solving

[1] Institute for Computing and Information Sciences, Radboud University Nijmegen, The Netherlands. Please direct correspondence to johank@cs.ru.nl.

http://dx.doi.org/10.7771/1932-6246.1141

Introduction

One of Leonhard Euler's most famous contributions to mathematics was his treatment of the Seven Bridges of Königsberg problem[1]. The city of Königsberg is set on both sides of the Pregel River. Both parts of the city, and two islands in the river, are connected using seven bridges. Euler was asked whether one was able to go for a stroll on a Sunday afternoon, passing each bridge exactly once and returning where one started the trip. Euler proved that the answer to this question was 'no'—no such tour exists (Euler, 1741).

Euler's main contribution, arguably, was not in the actual result, but in the way he tackled the problem. He quickly realized that the route one might take between the consecutive crossing of two bridges was irrelevant to solving the problem. The only relevant aspects of the problem are to be found in the topology of the bridges: which bridge connects which landmass. While solving the actual problem, Euler laid out the foundations of graph theory: in nowadays terms, each landmass corresponds to a vertex, and each bridge to an edge connecting vertices. The actual problem could be abstracted into a graph problem: given a graph G, does it contain an Euler tour, i.e., a tour connecting all vertices and traversing each edge in G exactly once?

However clever and thoughtful, Euler's treatment of the problem is quite typical of the way humans solve problems: by focusing on the relevant aspects of the problem only, dismissing details that are irrelevant. Nevertheless, we do make mistakes—some more than others. We sometimes include aspects that are not relevant (e.g., we might conclude that the distance between two bridges is relevant), leading to suboptimal representations; we sometimes weed out too many aspects (e.g., we might dismiss multiple bridges between two landmasses, focusing on a connectivity problem instead), leading to representations that may not lead to a correct solution.

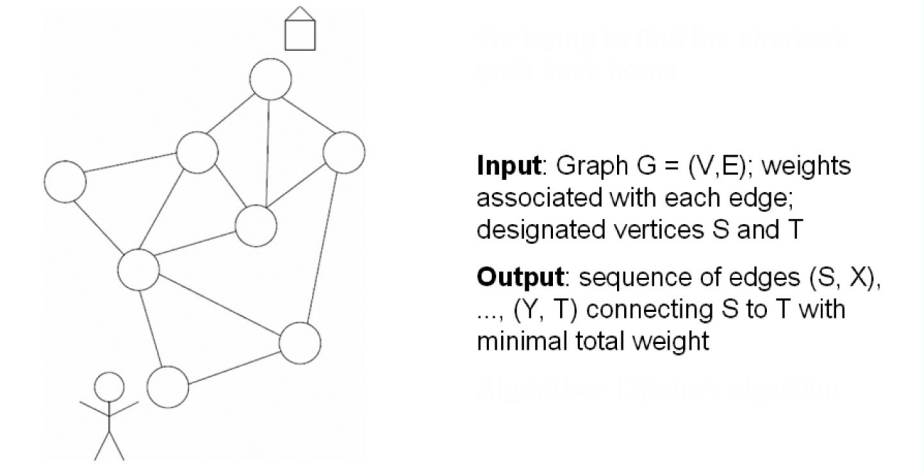

Figure 1. A problem and its formal description. Here the computational problem is to find a shortest path between two designated points S and T in a graph. It is formalized as an input-output mapping.

[1] http://en.wikipedia.org/wiki/Seven_Bridges_of_Königsberg

In general, it appears that the problem of finding a relevant abstraction significantly contributes to the overall complexity of solving the actual problem. Despite this observation, when studying a particular problem, a computational complexity analysis typically assumes that a relevant abstraction of the problem is readily available. For example, assume one wants to find one's way in an unknown city; in particular, one desires to travel from one's current location to the downtown hotel, using the shortest possible route. This 'real-world problem,' which we will denote Π_R, naturally translates into the computational problem Π_C of finding the shortest path in a graph with weighted edges and designated starting and ending points, as depicted in Figure 1. Such an abstract computational problem, cast into an input-output mapping, can then act as the starting point of a computational complexity analysis. We encode arbitrary instances of Π_C into input strings for a particular computational device (e.g., a Turing machine) and investigate whether the corresponding outputs can be computed efficiently, for example, in time, polynomial in the input size. In order for the problem to be tractably solvable, we require that the encoding is reasonable (we do not want to artificially increase the input size) and that the computation is feasible (Figure 2).

In this classical notion of computational complexity, however, we deal with abstract problems only, i.e., in order to assess the computational complexity of a particular problem, we do not take the abstraction step itself into account. However, there is increasing evidence from psychology (see, Ash, Cushen, & Wiley, 2009 for an overview) that finding a useful representation of a problem can be as hard as computing a solution to the (representation of the) problem. In order to construct and analyze computational models of

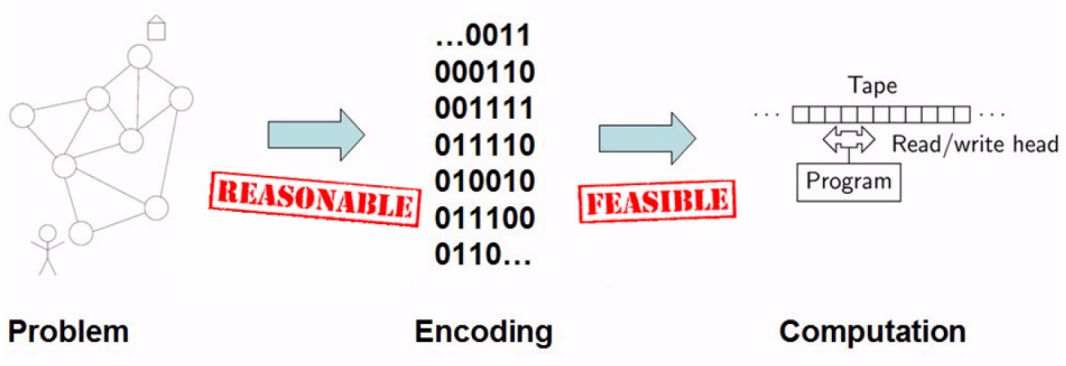

Figure 2. The classical computational complexity view on solving problems. A particular problem instance, expressible as a formal input-output mapping, is to be encoded in some computer-readable encoding; typically, but not necessarily, a string of binary numbers. The encoded instance is then fed as input to a computing device like a Turing Machine. In order for the problem to be solved tractably, we demand that the encoding is reasonable, i.e., that we do not 'blow up' the size of the instance, and that the computation is feasible, i.e., that the time needed to compute the output corresponding to a particular input, takes time, polynomial in the input size.

cognitive processes like problem solving (in a broad sense, including problems like intention recognition (Baker, Saxe, & Tenenbaum, 2009), visual perception (Cavanagh, 2011), analogy (Keane, 1988) and others), we need to address the issue of representation as well as the issue of computation.

This observation inspires an enhanced notion of computational complexity (Figure 3), where the abstraction step is explicit. The starting point of a complexity analysis here no longer is the abstract computational problem Π_C, but the real world problem Π_R: for this problem to be feasibly solvable, we not only require reasonable encodings and tractable computations, but also relevant abstractions. As an example, in the find-our-hotel problem, the distance between crossings is typically a relevant characteristic that should be included in the abstract computational problem. The color of the roof of the hotel is typically not relevant and thus should not be included in the abstraction. It will be clear that we can 'mess up' with the complexity analysis by inflating the problem instances, in a similar way as we can use unreasonable encodings, such that the resulting running time of the algorithm solving the problem—which is measured as a function on the input size—no longer reflects the actual difficulty of solving the problem.

In the remainder of this paper, we introduce a computational framework capturing the relevancy problem in Section 2. In Section 3 we discuss hard-to-abstract and easy-

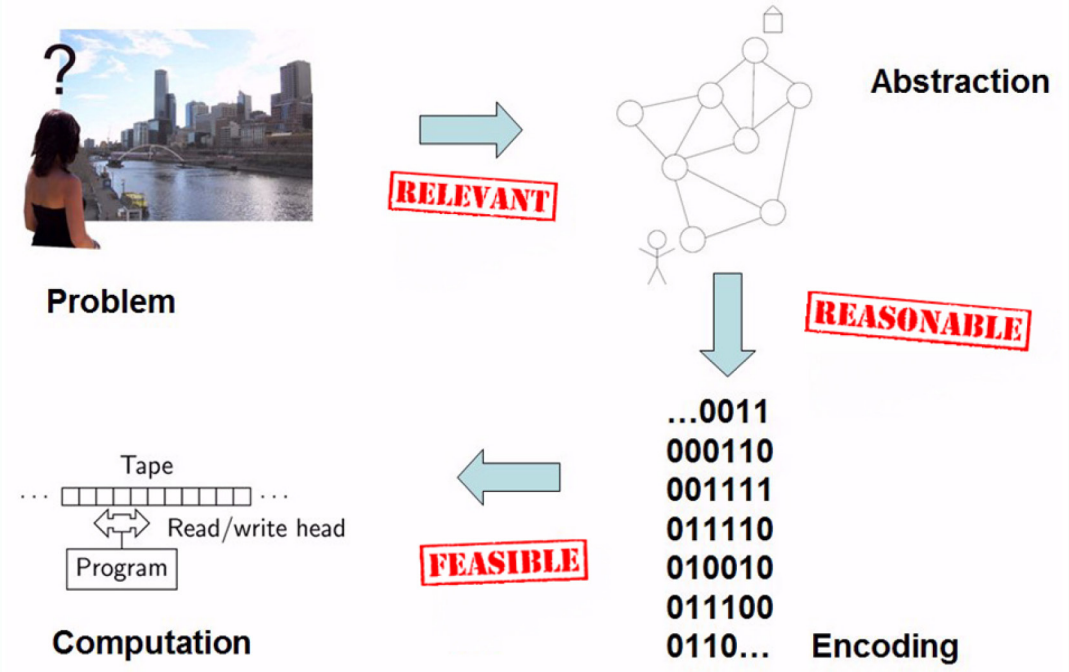

Figure 3. The enhanced computational complexity view on solving problems. Here we make an explicit distinction between the actual problem in the real world, i.e., finding one's way in an unknown city, and the abstract computational problem: finding a shortest path between two points. In addition to the previous demands on the encoding and the computation, we also demand that the abstraction captures the relevant aspects of the problem in the real world, e.g., distances between crossings, and abstracts away from many typically irrelevant details.

to-abstract problems, and in Section 4 the framework is put into context by discussing related research on representation, abstraction, and (insight) problem solving. In Section 5 we conclude and propose further research. A formal NP-hardness proof of the abstraction problem is given in the Appendix.

A Computational Framework for the Relevancy Problem

In the previous section we introduced an enhanced view on computational complexity analyses, which makes the abstraction step from the real world problem Π_R to abstract computational model Π_C explicit. Informally, this abstraction step seeks to isolate the relevant aspects of Π_R. In this section we introduce a computational framework capturing this abstraction step.

In the context of this paper, we see solving a problem Π_R simply as a transition from a problem-state to a solution-state, given some function f_Π mapping problems to solutions[2]. For example, the problem state might denote "Standing in front of King's Cross Railway station" and the solution state might denote "Walking on Trafalgar Square". Alternatively, these states might respectively denote "Sitting at the table, frowning and holding an unsolved nine-dot-puzzle" and "Standing up from the table, smiling and waving with a solved nine-dot-puzzle" or the problem and solution state may even refer to "X doesn't love me" and "X does love me," respectively. Note that the descriptions of these states are limited and can be extended ad infinitum, to encompass the entire state of the universe. When I am actually standing in front of King's Cross, I observe that there is a yellow car passing by, a business man is talking into his cell phone, the sun is shining and someone is just buying a newspaper at the local kiosk. Apart from everything that I observe through my senses, there is another infinite amount of information describing the current state, for example that it is raining in Uganda, that someone in Brooklyn just emptied her glass, and that the amount of radiation from the Pleiades is slightly above yesterday's level.

The bottom line here is that there are a massive amount of properties describing each state, yet hardly any of this information is relevant here. It is difficult to imagine a situation where the amount of radiation from the Pleiades influences the way I go from King's Cross to Trafalgar Square. In our problem solving, we abstract from all of these irrelevant properties immediately and focus on what is relevant, e.g., the underground train system connecting nearby stops, or the current time—if it's 3 a.m., we will probably need to take a taxi. Furthermore, while we assume that there is exactly one problem state (i.e., the state that I am currently in), many states qualify as a solution: me standing on Trafalgar Square with either bus 5, 11 or 15 riding by all are valid solution states.

[2] We are aware of the fact that this is just one notion of solving a problem. Other notions may include 'explain why *s* is a solution to $f_\Pi(p)$' or 'find a function f_Π matching *s* and *p*' or 'find a *p* that matches some given constraints'. In the context of this paper, we will assume that *s* and *p* are fully known and observable and leave other notions of solving a problem to future work.

We formalize "problem solving in the real world" as finding a transition from one point p in a state space (describing the initial problem state) to another point s in a solution region S of the state space, where the dimensions of the state space describe properties of p and s. Such a point corresponds to what McCarthy and Hayes (1969) call a situation: a complete state of the universe at a particular instant of time. Some of these dimensions may correspond to properties we typically regard as irrelevant, like "current amount of radiation from the Pleiades". Other dimensions correspond to obviously relevant properties, like "distance to the nearest underground station".

We assume that the number of dimensions N is massive, and that each dimension $d \in N$ can take an arbitrary (yet, for the sake of the formal analysis, bounded) number of values. Problem solving can then be formally defined as the transition from $p \in C^N$ to $s \in S \subseteq C^N$, where C denotes the maximal cardinality (or resolution) of the dimensions. The transition typically is in the form of some action plan, i.e., a set of intermediate states that allows us to move from p to s. Observe that N may—and will—be very large. Typically, only a small subset of N corresponds to relevant properties. We denote that subset with M, and assert that the actual transition from p to s is independent of values of dimensions outside M: whether the radiation from the Pleiades sums up to γ_1 or γ_2 will not influence my action plan, consisting of me taking the first available tube to Charing Cross. Our abstraction problem now is the problem of finding such a relevant subset $M \subseteq N$. We formalize the above informal notions into computational problems as follows. By the notation $x^{\downarrow M}$, we refer to the vector that is obtained from x by omitting all dimensions of x that are not in M. Furthermore, we assume that both p and s are fully observable and that every problem actually has a solution, i.e., S is non-empty.

Problem solving

Input: An input vector $p \in C^N$, denoting the problem-state; an output region $S \subseteq C^N$, denoting the solution-space, such that $f_\Pi(p) = S$ for a particular function f_Π mapping instances of a problem Π to solutions to these instances; an objective function o_Π such that $o_\Pi(s) = 1$ if $s \in S$, and $o_\Pi(s) = 0$ otherwise.
Output: A sequence of vectors $A_1, \ldots, A_k \in C^N$, defining an action plan to obtain $s \in S$ from p.

Abstraction

Input: A function f_Π mapping instances of a problem Π to solutions to these instances, with corresponding objective function o_Π.
Output: The smallest non-empty subset $M \subseteq N$ such that for every $s \in C^N$, $o_\Pi(s) = o_\Pi(s^{\downarrow M})$.

Without doubt, part of abstracting a real world problem into a computational problem is in *bringing down the resolution* of the problem dimensions (Palmer, 1978). For example, even when the distance to the nearest tube is a relevant dimension, we do not need to represent it in millimeters to construct action plans. However, we leave this aspect out of the abstraction problem, for reasons explained in Section 3.

Note that in the context of this paper we are not particularly interested in the action plans themselves, nor in how they are to be constructed or executed; our interest lies in the fundamental properties of the abstraction problem, i.e., the problem of obtaining the relevant characteristics of the original problem and omitting the non-relevant details. We do not focus on the question *how* we solve this abstraction problem, but rather on *what makes it hard*: we seek to find the *sources of complexity* of the abstraction problem.

In the next section we will discuss particular instances of the abstraction problem and discuss why some instances of the abstraction problem appear to be fairly easy, why others appear to be hard. We will introduce the notion of *expected relevancy* to help explain which dimensions are typically considered relevant, and others are considered irrelevant (or not considered at all), and how this notion can help in discriminating hard and easy to abstract problems.

Hard and Easy Abstraction Problems

For some problems, finding a good abstraction appears to be fairly easy. Most of us will agree that, for solving the problem of finding the shortest path to the downtown hotel, distances between crossings and landmarks are relevant characteristics of the problem that should be preserved in the abstraction, and the structure of the tiles in the pavement are irrelevant details that should be left out. It is not too hard to settle on a subset M of relevant dimensions in this problem. The problem generalizes well to similar problems, like finding the shortest path to the library.

Other problems are much harder to abstract. One such problem might be: "*Make X love me*" where the (presumed) begin-state p is: "*X doesn't love me*" and the desired end state is in the region S corresponding with "*X* does *love me.*" Here, there is for many dimensions d general disagreement whether $d \in M$ or not; arguably, there are many relevant dimensions. We will have a hard time explaining the relevant dimensions to others; in fact, it will be an educated guess at best. It is not well understood what the relevant dimensions are here. The problem does not well generalize to "*Make Z love me*", as typically the set of relevant dimensions will be different.

Are there problems that are *inherently* hard to abstract? In the context of the proposed formal framework, the answer can be an unambiguous '*yes*'. This need not surprise us, as the abstraction problem closely relates to the *frame problem* in artificial intelligence: the fundamental problem of deciding which information is relevant in order to make inferences, where in potential everything *might* be relevant[3] (Fodor, 1987). We show in

[3] There are many definitions of the frame problem; see, e.g., Haselager (1997) for an overview. Originally, the frame problem was restricted to the purely logical problem of deducing what stays the same as the result of an action (McCarthy & Hayes, 1969); we refer to a more general definition that sees the frame problem as a representational and inferential issue.

the Appendix, using a computational complexity analysis, that it is NP-hard[4] to decide in general that a particular dimension is relevant, and thus that no polynomial time algorithm can solve the **Abstraction** problem in polynomial time unless P = NP. Furthermore, we show that the (initial) resolution of the dimensions is in itself *not* a source of complexity: the **Abstraction** problem happens to be intractable (i.e., NP-hard), even if all dimensions have a resolution of *two*, that is, even if all dimensions in the real world problem would be binary (like yes/no, old/young, close/far). So, the intractability of finding a suitable abstraction cannot be ascribed to the task of bringing down the resolution of the dimensions to manageable sizes.

To discuss *why* the "*Make X love me*" problem is harder to abstract than "*Find shortest path to hotel*" we investigate when a dimension is typically considered (by the problem solver) to be irrelevant. It has been noted before (Kaplan & Simon, 1990, p. 403) that *"[We] are not equipped with generators for searching the space of 'all possible representations'"*. Hayes and Simon (1974) also demonstrate that, when confronted with a problem to solve, subjects don't *deliberately choose* among possible problem representations, but select the representation that is most obvious from the problem description at hand. Various studies on insight problem solving (e.g., Kaplan & Simon, 1990; Knöblich, Ohlsson, Haider, & Rhenius, 1999) suggest that we stick to a particular representation and we often need strong 'counter forces' (like hints) to reconsider what is and isn't relevant in solving the problem. These results suggest that we are typically able to quickly discriminate between features that are considered relevant or irrelevant (although we may be mistaken sometimes).

In the context of our formal framework, we introduce the notion of the **expected relevancy** δ_d of a dimension as some function mapping dimensions to the interval [0, 1], where 0 denotes *absolutely irrelevant* and 1 denotes *absolutely relevant*. We assert that the expected relevancy of a dimension is a subjective measure, conditioned on prior knowledge or experience, heuristics, and newly arriving information (such as hints from the experimenter, or consistent failure to solve the problem). Typically, the expected relevancy δ_d for a dimension *d* would be low if variation along the value of the dimension is not considered likely to affect the outcome of the problem solving activity. This notion of relevance corresponds to Wilson and Sperber's approach where "*an input is relevant to an individual when its processing in a context of available assumptions yields (. . .) a worthwhile difference to the individual's representation of the world*" (Wilson and Sperber, 2004, p. 608). It is also related to Gorayska and Lindsay's goal-oriented approach, where an aspect is relevant towards a goal if there is some action plan from the current state to the goal where this aspect plays an essential role (Gorayska & Lindsay, 1993). In a somewhat

[4] We assume that the reader is familiar with the notions P and NP for computational problems that can be decided, respectively for which a candidate solution can be verified, in polynomial time; the notion of NP-hardness, indicating presumed intractability of the computational problem, and the assumed inequality of P and NP. We refer the reader to, e.g., Garey and Johnson (1979) for a thorough discussion of these concepts.

different context, Müller, van Rooij, and Wareham (2009) proposed a related notion of a *relevant set of transformations* in similarity judging.

Note that the expected relevancy δ_d is a *subjective* measure that is typically *estimated* by the problem solver. Given this notion of expected relevancy, we suggest that one only includes dimensions in a abstraction that have a high expected relevancy. Naturally, making such an abstraction is easy when there are only few dimensions with a high expected relevancy, and hard when there are (many) dimensions with a high expected relevancy ("*Everything appears to be relevant!*") and when there are (many) dimensions for which the expected relevancy is difficult to estimate ("*I don't know what is relevant*!").

We should emphasize that we do not propose or even suggest that a problem solver actually consciously *computes* expected relevancies. Our cognitive processes often behave roughly according to the laws of probability theory, without the brain actually performing probabilistic inference before making a decision (Chater & Oaksford, 2008); similarly, abstractions can be made roughly according to the expected relevancy of the dimensions without (consciously) computing them. We *do* suggest that expected relevancy may be a measure that can be used to *describe* why some problems are harder to abstract than others. Likewise, we suggest that it may be a useful way of *explaining* individual differences in finding an abstraction, as well as explaining why people sometimes (like in insight problem solving) make unsuitable abstractions.

Discussion

The ability to solve problems is often considered to be one of the most complex intellectual abilities of mankind; it is therefore not surprising that studying (computational models of) problem solving is a key topic in cognitive science, perhaps with the seminal work of Newell and Simon (1972) as most impressive example. Even when we restrict ourselves to a narrow definition of problem solving (finding an action plan from a problem state to a solution state) there is a wild variety in the difficulty of problems to be solved. Naturally, some problems are more difficult than others. Funke (1991) contrasts simple and complex problems using criteria like the transparency of the problem and solution definitions, as well as complexity of the problem in terms of number and connectivity of variables. Complex problems may have intransparent definitions and large and highly connected problem spaces.

Here, some of the criteria for complexity appear to stem from *representing* the problem, while some others from *solving* the already represented problem. For some problems, it is both easy to find a suitable formal representation (i.e., a suitable abstraction of the real world problem) *and* solve the problem; finding the shortest path to the downtown hotel is such a problem. Finding a winning strategy in Go is fairly easy to cast into a formal

computational problem, but intractable to solve; other problems are easy to solve once we have a suitable representation.

As already noticed by Newell and Simon (1972), some problem solving processes are incremental in nature. They require only fairly routine or analytic steps to solve; the problem solver knows *how* the problem is to be solved, he or she just needs to apply the needed steps. While this may require in itself considerable intellectual effort, this type of problem solving can be characterized by a *steadily increasing Feeling-of-Warmth* (Metcalfe, 1986). Other problem solving processes, in contrast, involve non-incremental, discontinuous, or creative steps to "pass beyond a barrier". These problems are often denoted as *insight problems* (Smith, 1995; Sternberg & Davidson, 1995; Chu & MacGregor, 2011) as they typically require some insight or *aha-erlebnis* to solve. In such problems, Metcalfe (1986) found a *sharp increase* in Feeling-of-Warmth towards the end of the problem solving process. In these insight problems, the problem solver typically starts with an inappropriate representation, and the problem can only be solved when the problem solver realizes that the problem representation should be changed in order to 'break the barrier', for example (as in the Representational Change Theory) by relaxing some of the constraints (Ohlsson, 1992).

Kaplan and Simon (1990) mention a few perceptual cues in the Mutilated Checkerboard Problem[5] that help the participants to gain more insight in the problem, e.g., draw the attention on the (lack of) parity of the squares on the board. Apart from these 'external' forces (like hints and cues), they also mention more 'internal' means by which participants gained insight in the problem. In particular, they hypothesize that *noticing invariants* (features in the problem that do not change; e.g., in every partial covering two black squares remain uncovered) helps the problem solver in concluding that a re-representation of the problem is needed (after noticing that the color of the squares is a relevant feature). The reconsideration of features after either external or internal cues corresponds with the increase expected relevancy of the respective dimensions.

Insight problems thus are typically problems for which solutions are fairly easy to determine, once the correct representation is found. The correct representation is an appropriate abstraction that leaves out the irrelevant details but includes all the aspects of the problem that are relevant in finding a solution. An example is a variant of the so-called *Nine Dots Problem*. Assume you were given a sheet of paper, with nine black dots on it, ordered in a three-by-three square (Figure 4a), and are asked whether it is possible to connect the nine dots using only *three* connected straight lines. While the notion of problem solving here is slightly different as the solution itself is required (rather than a path to a given solution), the act of abstracting the problem is similar. We quickly (and likely, unconsciously) classify the way in which the dots are ordered as relevant for solving

[5] A problem in which participants are presented a checkerboard with two opposite squares (upper-right and lower-left, i.e., squares of the same color) removed and asked whether the board could be completely covered by domino pieces.

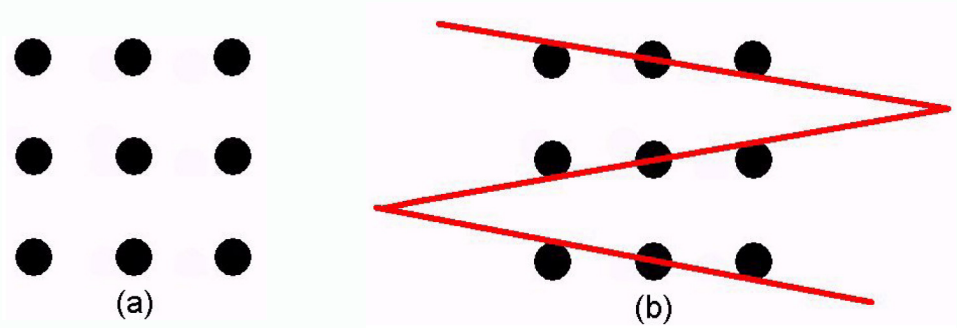

Figure 4. An example of the Nine-Dots-Puzzle (a) and a possible solution (b). Note that the solution depends on the dots being two-dimensional objects, rather than being dimensionless points in space.

the problem, and the thickness of the paper, the color of the dots, the exact placement of the dots on the paper, the current time, weather conditions, and NASDAQ-index as irrelevant. What makes this problem hard to solve is that we typically regard one aspect of the problem as irrelevant while it is in fact *crucial* in solving the problem, namely, the size of the dots. In typical mathematics-like problems, a 'dot' refers to a *point*, a zero-dimensional point in space. We abstract the problem by seeing it as nine *points* that need to be connected. In this particular problem, however, the dots *have* area, and one can connect them using three straight lines if we don't connect them at their center, but (like in Figure 4b) make use of the area of the dots. It is, however, fully explainable that we did not (initially) include the area of the dots as a relevant aspect that needed to be included in the representation, as it has a low expected relevancy.

As expected relevancy of a dimension is a subjective measure, it can also be used to explain individual differences in problem solving. Having domain-relevant expertise or knowledge typically influences the problem solving activity. This knowledge, however, can be a "*two-edged sword*" (Kaplan and Simon, 1990, p. 399) as it may both help and hinder the problem solver. It may help in discriminating between relevant and irrelevant aspects of the problem domain: novices may experience difficulties in assessing the expected relevancy of many dimensions. Experts, on the other hand, may bring in their past experience and knowledge to bring down the number of relevant dimensions to a manageable size. On the other hand, the expert may suffer from *functional fixedness* (Duncker, 1945)—the expert's prior knowledge (for example, "in mathematical problems, a *point* typically is dimensionless and has no area") may hinder his or her ability to solve the problem, as the expected relevancy of the dimension 'area of the dots' is low because of previous experiences. A novice that may not have any experience with mathematical puzzles may lack such a bias and may not immediately dismiss this dimension because of low expected relevancy. An extreme example of such an *expert blind spot* was given by Kaplan and Simon (1990, p. 379) where a very persistent graduate student in Chemical Engineering spent 18 hours on the Mutilated Checkerboard Problem, filling over 60 pages

in a notebook with all sorts of mathematical invariants, yet missing the parity of the color of the squares that are covered by a domino piece as a relevant dimension.

Conclusion

Whether a problem solving situation is trivial or complex depends on many factors; some of which can be classified as representational factors, others as computational factors. The *computational complexity* of (an abstraction of) a problem is a well-studied area, with its own conferences, research groups, and the famous $ 1,000,000 question[6] whether the classes P and NP are equal or different. The *representational complexity* of a problem—how hard is it to find a suitable formal representation of a problem—has received much less attention from computer science. In this paper we proposed one possible computational framework for studying the representational complexity of a problem and related it to intuitive notions of hard and easy to abstract problems using the *expected relevancy* measure. We proposed that humans experience difficulty in finding a suitable abstraction if there are many features with a high expected relevancy, or when the expected relevancy of many features is difficult to assess. This notion of expected relevancy was then used to explain why people experience difficulties solving insight problems like the Nine Dot Problem. Furthermore, this notion was used to explain both the possible help and hindrance of previous experience or knowledge.

While the expected relevancy is a *subjective* measure on a problem aspect, it can be shown that, within this framework, there exist problems that are *inherently* hard to abstract[7]. In future work it would be very interesting if these hard problems can be characterized in some way such that they can be formally separated from the easy problems. Furthermore, while it is NP-hard to find a *minimal* subset of relevant aspects, we do not have formal results yet on approximation (finding a subset that is guaranteed to be *almost* minimal) or randomization (finding a subset that is minimal with a *high probability*, but may be way off in extreme cases) strategies of abstraction. Lastly, we did not address the question *how* people[8] assess the expected relevancy of a dimension; we believe it is beyond the scope of this paper (and the author's expertise) to make any claims on this subject.

---

[6] The Clay Mathematics Institute has denoted this open problem as one of the "Millennium Prize" problems and offers a $ 1,000,000 prize for anyone proving either P = NP or (much more likely) P ≠ NP.

[7] This may not be surprising given similar results for the Feature Selection problem in machine learning (Koller & Sahami, 1996; John, Kohavi, & Pfleger, 1994; Bin, Jiarong, & Yadong, 1997).

[8] Or, for that matter, an artificial intelligence. Note that we did not suggest that expected relevancies were to be computed for every possible dimension—that would lead to a frame problem 'in disguise'.

Acknowledgments

This paper is based on the presentation that the author delivered at the Dagstuhl Seminar on Computer Science & Problem Solving: New Foundations. I am indebted to the participants of this seminar for valuable feedback and suggestions. In particular, I wish to thank Tom Heskes, Iris van Rooij, Todd Wareham, Stefan Leijnen, Mark Blokpoel, and two anonymous reviewers for helpful discussions, literature suggestions, and feedback on earlier versions of this paper, and Bas Maes for designing the figures in the paper.

References

Ash, I. K., Cushen, P. J., & Wiley, J. (2009). Obstacles in investigating the role of restructuring in insightful problem solving. *Journal of Problem Solving, 2*(2), 6–41.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition, 113*, 329–349. http://dx.doi.org/10.1016/j.cognition.2009.07.005

Bin, C., Jiarong, H., & Yadong, W. (1997). The minimum feature subset selection problem. *Journal of Computer Science & Technology, 12*(2), 145–153.

Cavanagh, P. (2011). Visual cognition. *Vision Research, 51*(13), 2011, 1538–1551.

Chater, N., & Oaksford, M. (2008). *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford University Press, Cary, NC. http://dx.doi.org/10.1093/acprof:oso/9780199216093.001.0001

Chu, Y., & MacGregor, J. N. (2011). Human performance on insight problem solving: A review. *Journal of Problem Solving, 3*(2), 119–150.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, *58*(5), Whole No. 270).

Euler, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae, 8*, 128–140. Retrieved from http://www.math.dartmouth.edu/~euler/pages/E053.html.

Fodor, J. A. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Z.W. Pylyshyn (Ed.), *The robot's dilemma: the frame problem in artificial intelligence*. Norwood, NJ: Ablex.

Funke, J. (1991). Solving complex problems: Exploration and control of complex social systems. In R. J. Sternberg and P. A. Frensch (Eds.): *Complex problem solving: Principles and mechanisms*. Hillsday, NJ: Lawrence Erlbaum Assocates.

Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness*. San Francisco, CA: W.H. Freeman and Co.

Gorayska, B., & Lindsay, R. (1993). The roots of relevance. *Journal of Pragmatics*, *19*(4), 301–323. http://dx.doi.org/10.1016/0378-2166(93)90091-3

Haselager, W. F. G. (1997). *Cognitive science and folk psychology: The right frame of mind.* London: Sage.

Hayes, J. R., & Simon, H. A. (1974). Understanding written problem instructions. In L. Gregg (Ed.), *Knowledge and cognition*. Potomac, MD: Lawrence Erlbaum Associates.

John, G. H., Kohavi, R., & Pfleger, K. (1994). Irrelevant features and the subset selection problem. *Proceedings of the Eleventh International Conference on Machine Learning,* pp. 121–129.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22*, 374–419. http://dx.doi.org/10.1016/0010-0285(90)90008-R

Keane, M. T. (1988). *Analogical problem solving*. Oxford, UK: Halsted Press.

Knöblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1534–1555. http://dx.doi.org/10.1037/0278-7393.25.6.1534

Koller, D., & Sahami, M. (1996). Toward optimal feature selection. *Proceedings of the Thirteenth International Conference on Machine Learning*, pp. 284–292.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65–100. http://dx.doi.org/10.1111/j.1551-6708.1987.tb00863.x

McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence (Vol. 4)*. Edinburgh, UK: Edinburgh University Press.

Metcalfe, J. (1986). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*(4), 623–634. http://dx.doi.org/10.1037/0278-7393.12.4.623

Müller, M., van Rooij, I., & Wareham, T. (2009). Similarity as tractable transformation. *Proceedings of the 31st Annual Conference of the Cognitive Science Society,* pp. 50–55.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. Keane & K. Gilhooly (Eds.), *Advances in the Psychology of Thinking* (pp 1–44). Hemel Hempstead, UK: Harvester-Wheatsheaf.

Palmer, S. E. (1978). Fundamental aspects of cognitive representation. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization*, 259–303. Hillsday, NJ: Lawrence Elbaum Associates.

Simon, H. A. (1978). On the forms of mental representation. In C. W. Savage (Ed.), *Perception and Cognition: Issues in the Foundations of Psychology* (pp. 3–18). Minneapolis, MN: University of Minnesota.

Smith, S. M. (1995). Fixation, incubation, and insight in memory, problem solving, and creativity. In S. M. Smith, T. B. Ward, & R. A. Finke (Eds.), *The creative cognition approach* (pp. 135–155). MIT Press, Cambridge, MA.

Sternberg, R. J., & Davidson, J. E. (Eds.) (1995). *The nature of insight*. Cambridge, MA: MIT Press.

Wilson, D., & Sperber, D. (2004). Relevance theory. In Horn, L. R. & Ward, G. (Eds.) *The handbook of pragmatics* (pp. 607–632.). Oxford, UK: Blackwell.

Appendix

In the Appendix, we argue that it is in general intractable to solve the **Abstraction** problem defined in Section 3, that is, to find the smallest relevant subset M of the set of all dimensions N. In particular, we show that it is NP-hard in general to decide whether a particular dimension is relevant. We start by defining two computational decision problems, and show that they are encapsulated in the **Abstraction** problem mentioned above. Then we proceed to prove NP-completeness, respectively co-NP-completeness, of these decision problems, thus establishing the above mentioned results. We assume that the reader is familiar with the complexity classes P, NP and co-NP, and with intractability proofs in general. We refer to textbooks like Garey and Johnson (1979) for more background.

We call a variable x_i *redundant* in a Boolean formula φ if x_i's truth assignment does not influence the characteristic function[9] $\mathbf{1}_\varphi$ of φ. In other words, x_i is redundant in φ if $\mathbf{1}_\varphi(x_i = 0) = \mathbf{1}_\varphi(x_i = 1)$. In contrast, x_i is *relevant* in φ if $\mathbf{1}_\varphi(x_i = 0) \neq \mathbf{1}_\varphi(x_i = 1)$, that is, if x_i's truth assignment *does* influence the characteristic function. These notions induce the following decision problems.

isa-Relevant Variable

Input: Boolean formula φ with n variables, describing the characteristic function $\mathbf{1}_\varphi: \{0,1\}^n \rightarrow \{1, 0\}$, designated variable $x_i \in \varphi$.
Question: Is x_i a relevant variable in φ, that is, is $\mathbf{1}_\varphi(x_i = 1) \neq \mathbf{1}_\varphi(x_i = 0)$?

isa-Redundant Variable

Input: Boolean formula φ with n variables, describing the characteristic function $\mathbf{1}_\varphi: \{0,1\}^n \rightarrow \{1, 0\}$, designated variable $x_i \in \varphi$.
Question: Is x_i a redundant variable in φ, that is, is $\mathbf{1}_\varphi(x_i = 1) = \mathbf{1}_\varphi(x_i = 0)$?

It can be readily shown that these problems are closely related to the **Abstraction** problem: every variable in φ translates to a (binary valued) dimension in C^N with d as designated dimension corresponding to x_i; f_Π outputs a state corresponding to a satisfying truth instantiation; the characteristic function $\mathbf{1}_\varphi$ then corresponds with the objective function o_Π. We denote the problem state p as an arbitrary vector in this space and the solution space S as a region in C^N corresponding with the satisfying truth instantiations. The smallest subset M of N such that $o_\Pi(s) = o_\Pi(s^{\downarrow M})$ now contains only dimensions that correspond to relevant variables; any other dimension $d' \in N \setminus M$ corresponds to a re-

[9] The characteristic function of a Boolean formula φ, denoted by $\mathbf{1}_\varphi$, maps truth assignments to φ to {0, 1}, such that $\mathbf{1}_\varphi(\mathbf{x}) = 1$ iff. $\mathbf{x}$ denotes a satisfying truth assignment, and 0 otherwise.

dundant variable. Naturally, any algorithm that solves **Abstraction** can also be used to decide **isa-Relevant Variable** and **isa-Redundant Variable**: from φ, construct an **Abstraction** instance as described above and compute M; x_i is relevant if and only if $x_i \in M$, otherwise x_i is redundant. As $M \subseteq N$, this computation can be done in polynomial time. As we will show shortly that deciding **isa-Relevant Variable** and **isa-Redundant Variable** is NP-complete, respectively co-NP-complete, it follows that computing the relevant subset of dimensions (or deciding whether a particular dimension is relevant) for f_Π is intractable, unless P = NP.

Theorem: **isa-Relevant Variable** and **isa-Redundant Variable** are NP-complete, respectively co-NP-complete.

Proof. To prove NP-completeness, we reduce **isa-Relevant Variable** from the well-known NP-complete **Satisfiability** problem (given a Boolean formula φ with n variables; is φ satisfiable?). Membership of NP follows as we can verify in polynomial time a certificate, consisting of a tuple of two truth assignments ($\boldsymbol{x}, \boldsymbol{x'}$) such that $\mathbf{1}_\varphi(\boldsymbol{x}) \neq \mathbf{1}_\varphi(\boldsymbol{x'})$. NP-hardness is proven as follows. Let ψ be an instance of **Satisfiability** with n variables and let $\varphi = \psi \wedge x_{n+1}$.

$\rightarrow$ If x_{n+1} is a relevant variable, then ψ is satisfiable, as there exists by definition at least one truth assignment to φ such that $\mathbf{1}_\varphi(\psi \wedge x_{n+1} = \text{true}) \neq \mathbf{1}_\varphi(\psi \wedge x_{n+1} = \text{false})$, and given that $\psi \wedge x_{n+1} = \text{false}$ always evaluates to false, we have that ψ is necessarily satisfiable.

$\leftarrow$ if ψ is satisfiable, then x_{n+1} is a relevant variable as $\mathbf{1}_\varphi(\psi \wedge x_{n+1} = \text{true}) \neq \mathbf{1}_\varphi(\psi \wedge x_{n+1} = \text{false})$ for any satisfiable truth instantiation to ψ, given that $\psi \wedge x_{n+1} = \text{false}$ always evaluates to false.

As we can obviously construct the above reduction in polynomial time, that proves that **isa-Relevant Variable** is NP-complete. The co-NP-completeness proof of **isa-Redundant Variable** is almost similar, save that we reduce from the co-NP-complete problem **Tautology** (given a Boolean formula φ with n variables; is φ a contradiction?) and we prove membership in co-NP by providing the same certificate as above, but now used as a counterexample.

(Q.E.D.)