Lund University Cognitive Science

Beyond Accuracy: How Models of Decision Making Compare to Human Decision Making

Master Thesis in Cognitive Science
Lund University
Sweden
June 2005

Author: **Carl Christian Rolf**
810822-3952

Supervisor: **Annika Wallin**

Abstract

Decision aids are becoming more popular and more accessible thanks to the internet and hand held computers. The question remains, however, if decision aids can be used beyond just aiding the human decision process. Can we use computers to make decisions without loosing out on some of the aspects of decision making that are important to us? The aspects that will be dealt with are; transparency, consistency, accuracy, improvement, adaptability, and speed.

In this master thesis I will present what these aspects mean to decisions, and how different models of decision making compare to human decision making with regard to them. For instance; how transparent and accurate are humans in their decision making compared to linear models, designed to fit the available data? I will also give some examples of decision aids and how they handle the features that define a good decision.

Introduction 3
Definitions 4
The Models 5
  Linear Models 5
  Bayesian Networks 6
  Take The Best 6
Makings of a Good Decision 7
  Transparency 8
    Linear Models 9
    Bayesian Networks 9
    Take The Best 10
    Human Decision Making 10
  Consistency 11
    Linear Models 12
    Bayesian Networks 12
    Take The Best 13
    Human Decision Making 13
  Accuracy 13
    Linear Models 14
    Bayesian Networks 14
    Take The Best 15
    Human Decision Making 15
  Improvement 16
    Linear Models 16
    Bayesian Networks 17
    Take The Best 17
    Human Decision Making 17
  Adaptability 17
    Linear Models 18
    Bayesian Networks 18
    Take The Best 18
    Human Decision Making 19
  Speed 19
    Linear Models 20
    Bayesian Networks 20
    Take The Best 21

Human Decision Making 21
Summary 21
Relations between Aspects 22
Transparency 22
Consistency 22
Accuracy and Speed 23
Improvement 23
Adaptability 23
The Models from a Search Tree Perspective 24
Linear Models 24
Bayesian Networks 26
Take The Best 26
Summary 27
Examples of Decision Aids 27
Sperling's Best Places 28
GeNIe 29
Fast and Frugal Decision Trees 31
Summary 32
Discussion 32
Weighing the Aspects 33
Which Decision Aids are best? 35
Should we use Decision Aids? 35
Conclusions 36
References 37

Introduction

This thesis will try to clarify what the different aspects of a good decision are and how they are related. It will also examine the performance of models and human decision making with regard to these aspects. The goal is to show that decision aids could improve not just the accuracy of expert decisions, but also such aspects as transparency. This is an important step in motivating the use of decision aids, since it answers the question: Can we use decision aids, without loosing performance in some of the aspects of good decision making? There are several aspects that define good decisions; transparency, consistency, accuracy, improvement, adaptability, and speed. The transparency of a decision means that an outside observer can identify how the decision was reached. Consistency means that given the same, or very similar, information, the same decision will be reached. Accuracy is how correct decisions are on average. Improvement means that the decisions improve over time when that is possible. Adaptability means that decisions can be adjusted to fit the current policy. Speed is how long, in a suitable unit, it takes to make the decision. Although there is no formal requirement for a good decision to exhibit good performance with regard to all these aspect, it is nonetheless something that should be striven for.

When decision making is studied or modeled, accuracy is usually the authors' main concern. Rarely have comparisons of different models included performance with regard to both speed and accuracy. When designing decision aids intended for actual use, however, the aspects mentioned above become important. These aspects are not only important when comparing models to each other, but also when motivating the use of decision aids to the general public. Transparency, for instance, is of importance since, if no one could tell how or why a decision was made, there is no way to improve the decision without starting from scratch. This is one of the reasons for why judges in a court of law normally present the reasoning behind their decision. In the case of courts consistency is also an important aspect; given the same set of circumstances they defendant should be judged the same way. The issue of judges' sentencing consistency has recently (January 2005) come into focus since the US Supreme Court ruled that the mandatory sentences set by Congress should merely be considered as guidelines. These mandatory minimum punishments had been set up by congress to achieve a higher level of consistency between the different states. Presumably this reinterpretation will lead to a decreased consistency.

Computer-based decision aids, i.e., devices that will give you recommendations on what decision to make as you input the information you have at hand, all rely on mathematical models. Some of them, like Bayesian networks, are highly complex, while others, like Take The Best, are simple in execution but rely on quite heavy pre-computations (Jensen, 1996; Gigerenzer & Goldstein, 1996). What they all have in common is that they reduce any given situation to mere numbers. Such a reduction is a quite controversial issue. This will be discussed later, and as the reader will see it may not be such a big problem after all.

The most common decision aids available today are web-based tools using weighed averaging to sort the different alternatives presented to the user. An example of such a decision aid is *Sperling's Best Places* (http://www.bestplaces.net), that lets you input what kind of factors you value most when selecting a place to live. The alternatives are then sorted based on the weights that you gave. Since such decision aids are designed to sort similar options they are usually restricted to a specific field, such as selecting a good city to live in.

It is doubtful whether one can create a general decision aid; all models that rely on statistical data need that data to be specific in order to have a high accuracy. Another drawback is that they often employ a non-compensatory winnowing process, i.e., a process that removes alternatives that seem unlikely to rank high. Such winnowing might remove alternatives that would have won if the calculations of the weighed averages had included those (Edwards & Fasolo, 2001).

The models that will be studied in this master thesis are: linear models, Bayesian networks, and Take The Best (henceforth: TTB). Linear models operate in the same weighing fashion as described above, and usually rely on multiple regression to attain the optimal weights. Bayesian networks utilize Bayes' theorem in a causally connected network to propagate the information. TTB relies on "ecological validity", the relative frequency of correct predictions, to order the information, and then bases its advice on the first cue that differentiates between the available advice. Human decision making process will also be analyzed, since that is the decision process that would be replaced by the use of decision aids. For more detailed descriptions of the different models see Brehmer (1988) for linear models; Jensen (1996) for Bayesian networks; and Gigerenzer and Goldstein (1996) for TTB.

Definitions

Since it is easy to confuse different kinds of decisions here are the definitions that will be used throughout the thesis.

Advice – the course of action recommended by a decision aid.
Advice process – the process by which a decision aid chooses the advice to give.
Aided decision – a decision made by a human with the help of a decision aid.
Cues – information elements about the world, e.g., eye color, blood pressure.
Decision maker – the human making the decision.
Decision – a course of action, determined with or without the use of a decision aid.
Decision aid – something, or someone, giving advice on what course of action to take.

Decision process – the act of reaching an unaided decision.
Options – the possible decisions or advice in the given situation.
Outcome – the effects and consequences that follow because of the decision.
Unaided decision – a decision made by a human without the use of a decision aid.

In the case of a doctor seeing a patient, the doctor follows a decision process, and then makes an unaided decision to give a certain diagnosis. If, given the information available, several different diagnoses were possible these were the doctor's options. If the diagnosis was incorrect one of the outcomes is that the patient receives incorrect treatment. If the doctor had been using a decision aid, its advice had been on what diagnosis to make. Provided that the doctor did not completely disregard the advice he would then have made an aided decision. In this situation it is assumed that the diagnosis only has one treatment, otherwise the patient would also be involved in the decision process.

The Models

The models chosen in this thesis are the most well known models of decision making. They are all *actuarial* models, i.e., they are based on statistical calculations. While linear models and Bayesian networks were developed as a means of mathematical decision making, TTB was developed in an attempt to model human rationality. An example of human rationality modeled by TTB is the selection task of identifying which of two cities is the largest. The main difference between these approaches is that models of human rationality sometimes take the approach of *bounded rationality* (Simon, 1955). This means that one tries to minimize the memory accesses and calculations required in the decision process. Linear models and Bayesian networks in contrast were not designed with any constraints of memory or calculation power in mind.

Linear Models

Linear models stem from the *lens model equation* (Hursch et al., 1964), which is based on Brunswik's lens model of human perception (Brunswik, 1956).

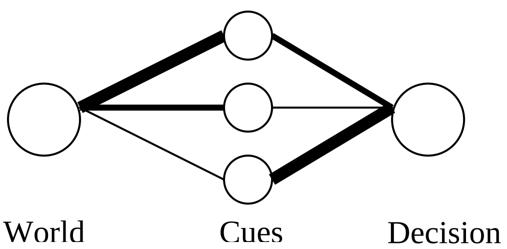

Figure 1: Example of a lens model. The weighing is indicated by the line width.

The general idea is that there are cues from the world that are weighed and then used in the decision process. For example, when diagnosing a disease one gathers information about blood pressure and body temperature, then weigh these before calculating what diagnosis to give. By determining these weights through multiple regression analysis of previous situations one will get the optimal weights from the specific situation. The downside, as we shall se later, is that the model becomes more difficult to use the more alternative decisions there are.

In order to separate these decision one will need more cues and therefore need more information when making the decision.

Bayesian Networks

Bayesian networks differentiate from linear models mainly in that the statistical analysis is not undirected. Multiple regression is undirected in the sense that there are no presupposed dependencies between the cues. This means that one has to test many more situations in order to extract the most important cues. In a Bayesian network, however, the dependencies are given by the structure of the network. These dependencies are usually constructed from *a priori* information about the situation. Another approach is to generate the network directly from the data, which is a computationally intensive procedure.

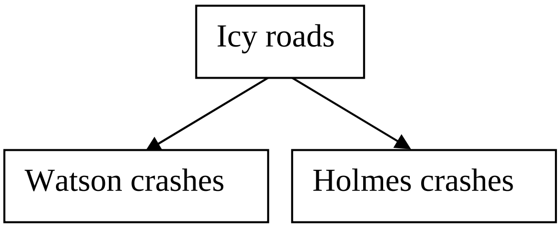

Figure 2: Bayesian network for if Holmes or Watson crashes their car given icy roads.

Figure 2 shows a Bayesian network, in each node there are two probabilities; “yes” and “no”. The probability of the road being icy and the probabilities that Holmes or Watson will crash when the road is icy, and when it is not, are given. With this information it is possible to calculate the probability that the roads are icy if we find out that Holmes or Watson has crashed. Since Bayesian networks portray the causality of the situation being modeled, they are often referred to as causal networks. For a study on this interpretation of Bayesian networks see Pearl (2000).

There are several algorithms for propagating the information in the network; in the worst case they are all NP-hard. The newer algorithms use mathematical tricks, such as triangulating the network graph and finding cliques of nodes to generate a junction tree (Jensen, 1996). Junction trees are trees with specific mathematical properties, which can be used to optimize the propagation of information.

Take The Best

Take The Best is a non linear model that uses a simpler form of statistical analysis. Instead of performing a linear regression, it simply looks at the relative frequency with which the cues predict the outcome. Then it runs through these cues serially, starting with the most relevant cue, i.e., the cue with the highest relative frequency. This frequency is called cue validity or *ecological validity*. Another measurement is the discrimination rate, which is a measurement of how good the cue is at separating two alternatives. If the cue that is being looked at cannot separate the two alternatives, the next cue is looked at. This is a non-compensatory procedure; a cue that did not separate the alternatives is not taken into account at all later in the algorithm.

Since TTB often returns an advice before it has checked all the cues, it needs less memory access than a linear model. Therefore it is said to be “fast and frugal”, meaning that it is faster than looking at all the cues, without loosing accuracy compared to other models. It should be noted, however, that most experiments involving TTB have been done using simple selection tasks, such as identifying which of two German cities is largest (Gigerenzer et al., 1999).

In contrast to the other models, TTB was not really designed as a decision aid. Given its accuracy in the experiments mentioned, and the large amount of publications about fast and frugal heuristics, however, it will be included for comparison. Since there are currently no decision aids available that use TTB, fast and frugal decision trees will be used later in the example section as the instance of how a decision aid based on a fast and frugal model works in reality. Fast and frugal decision trees are decision trees that are generated by using a heuristic similar to TTB. The reason TTB, and not fast and frugal trees, will be studied in the main part of this master thesis is that such trees are not designed as much to aid decisions as guide the human decision process.

Makings of a Good Decision

As mentioned in the introduction there are several aspects that define good decisions. Often, however, the performance of a model is measured only by its accuracy. With fast and frugal heuristics, speed is also a factor. Such narrow views of what performance means, however, are not sufficient when trying to motivate a more widespread usage of decision aids. Most of the criticism against using models of decision making in, e.g., medical diagnosis is not about lack of accuracy, but rather lack of “humanity”. This term usually refers to the need to reduce every situation to pure numbers when using a decision aid. Such a reduction seems intuitively wrong, often, however, numbers such as grades have been given by other humans. Breaking down good decisions into their components makes it easier to argue against such vague terms as “humanity”. The lack of usage of decision aids, despite the high accuracy of, e.g., linear models, clearly shows the need for a wider interpretation of performance. If decision aids outperform human decision making in every aspect of a good decision, it is hard to argue against using them.

The aspects mentioned above are not quite independent, and they interact at different levels. For instance, improvement of accuracy leads to reduced consistency since the new decisions will be better, and therefore different from the previous ones. However, this particular loss of consistency can be avoided by redefining the period for which consistency is measured when better data becomes available. The aspects also differ in importance. As mentioned above, transparency is important in court decisions, but in some decisions, that need to be made on the fly, the importance of transparency can be insignificant compared to accuracy or speed.

There are also cultural- and policy-related differences in what weights different aspects receive. To take an example; society being tougher on crime means a reduction of consistency with previous decisions. This means that the policy on how to treat the different outcomes of a decision play a major role in defining what makes a good decision good. When considering different outcomes it seems that politicians and other policy-setters miss the fact that reducing only one type of error increases the other (Hammond, 1996). This is best explained through a Taylor-Russell diagram (Taylor & Russell, 1939).

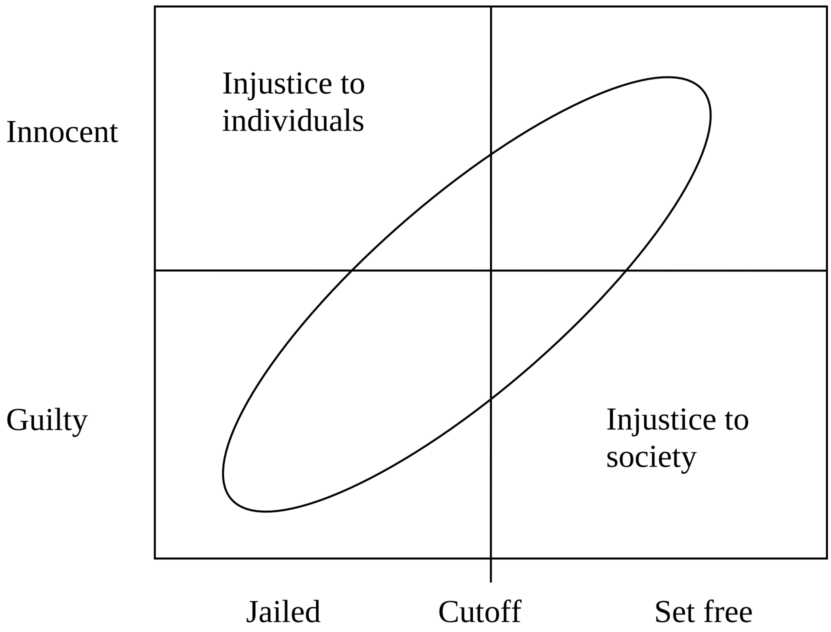

Figure 3: Taylor-Russell diagram of a court decision.

Injustice to individuals in this case is an example of a false positive, the injustice to society is an example of a false negative. Shifting the cutoff to reduce one kind of injustice inevitably increases the other kind. Only by reducing the area of the outcome ellipse, i.e., making more accurate decisions, can both injustices be reduced. This is why using decision aids to improve accuracy is important. However, it is also important that the decision aid can be made to reflect the current policy.

Transparency

There are many decisions that need to be transparent; court rulings and major decisions by politicians need to be open for journalists and voters to scrutinize. Without transparency in the decision processes of figures of authority there would be no way to assess whether the power given to the decision makers is well deserved, or if it needs to be restricted or removed.

There is, however, another reason for transparency, namely feedback. If the decision maker cannot go back and investigate why he did what he did, there is no way to know what went wrong or what he did right. To benefit learning, feedback should also be given in the form of differences between how the decision was reached and how it should have been made to be more correct. An example of this can be to give the difference between a decision maker's weighing of cues and the weights of the optimal linear model. Decision aids that are based on models that are more similar to how humans make decisions can be better from a transparency perspective since they can provide more intuitively formulated feedback. Lack of proper feedback is all too common for people making important decisions. The reason is that we do not know the outcome of the decision we did not make, leaving us with information about only one outcome, and even then we often do not know which category in the Taylor-Russell diagram it belongs to. For instance, a doctor giving an incorrect diagnosis might have to wait several years before the correct diagnosis becomes apparent. As expected, a lot of research, e.g., in medicine shows that experts hardly learn at all from experience if they do not have access to feedback on how to make correct decisions (Wigton et al., 1986; Brehmer, 1980; Camerer & Johnson, 1991; Dawes, 1994).

There are two separate kinds of transparency; transparency of how the model was designed, and transparency of how the model reached the advice it gave. In models that require pre-calculations it may be of interest to know how these were chosen and performed. For instance, it might be of interest to see how a Bayesian network was inferred from the data available. In this thesis, however, I will only look at the transparency of how the advice was reached, and not how the advice process was designed.

Linear Models

The transparency of linear models is, as with most computer models, excellent in principle. Given some mathematical understanding, it is easy to see how the different cues are weighted and how this steers the advice process in a certain direction. One can also identify whether there are any small differences that would have led to a different advice. For instance, one could see what happens if some cues are set as unknown, or what would happen if the most important cue is given less weight.

Giving feedback through linear models is quite easy if one has a good idea of the cues used by the decision maker. One can calculate the weights that the decision maker had for his unaided decision and compare them to the optimal weights. This can be used to show what information should be taken into more (or less) consideration. This is what was done in the study by Wigton et al. (1986). In this study medical students were given the task of diagnosing urinary tract infection. One group was given simple feedback, telling them only if they had been correct or not. The other group was given feedback in the form of the weights that the optimal linear model had, and how their weighing compared to it. The student's weights were derived by calculating the weights that a linear model would have if it had reached the same decisions as the student. The group receiving simple feedback hardly improved their accuracy at all, while the other group improved significantly. The problem with feedback through the weights of cues is that one makes the assumption that the decision maker used all the cues that the model uses. However, since cues are usually recommended by experts in the specific field, there is hopefully a strong correlation in these domains. Unfortunately the practice of giving feedback in the form of weights is quite rare in most areas of expert decision making (Dawes, 1988).

Although linear models are easy enough to understand, there is one complication, namely that one must use first order relations when designing them. This means that if the connection between a cue and its effect is, e.g., of the second order, one must reformulate the cue so that the relation is of the first order. For example, both a high and a low blood pressure are bad for your health. Therefore a cue for blood pressure has to be formulated in the form of "distance to optimal blood pressure". This may complicate the understanding not only of how cues were chosen but also of how they relate to one another. It is rare that all variables used in the decision process are independent.

Bayesian Networks

The calculations performed in a Bayesian network are quite a bit more complex than those in linear models or TTB. Therefore it is harder for the average person to understand how the advice process operates. However, the structure of the network, i.e., the order of the nodes and the connections between them, provide a good way of understanding the causality of the situation. This means that one does not have to understand the underlying math to understand why the network looks like it does (a design principle which is often good in computer software engineering).

Despite the more tangible representation of how the information propagates, it is often harder to see how different information would have lead to different advice. The reason is that the information usually has to pass through several nodes before reaching the advice node. This means that the information will be filtered through several layers of probability calculations. This works against people's desire to find a single cue to directly base their decision on. That desire often becomes apparent when looking at the models people build when trying to predict economic development (Hammond, 1996). Since Bayesian networks do not strive for simplicity, but thrive through complexity they may seem highly unintuitive to the average user.

Presenting feedback in the form of Bayesian networks is an interesting possibility. By doing so, one can present a well structured model of familiar decisions, such as a diagnosis. Letting the decision maker experiment with probabilities and information in the network should be more educational if the setting and decision is familiar. However, this kind of feedback would not work in the same manner as presenting the weights of a linear model. The feedback would be in the form of the relations between cues rather than their relative importance. Making the causality more apparent would probably make it easier for the decision maker to observe the decision process from an outside perspective. This gives her an opportunity to analyze and improve her decisions.

Take The Best

Just like linear models it is easy to se how advice was reached using TTB. If one wants to estimate which advice will be given, however, there are problems. Since TTB looks at one cue at a time and bases its advice on that single cue, there is no way to predict, beyond simple frequency, in which direction the advice process will go in the next step. For example, if all cues but one lead to advice A, but the cue leading to advice B has the highest ecological validity, advice B is given if the first cue discriminates. Despite this lack of predictability during the advice process, TTB is fully open to investigation into how the advice was finally reached.

It is more difficult to use TTB for feedback than it is to use linear models. This is because it is harder to tell whether the same cues were used by the decision aid and the decision maker, or if the cue validities were incorrect. This means that even though TTB is possibly a more correct model of human judgment than, e.g., linear models, it is easier to improve the decision maker's accuracy if one uses models more suitable for feedback than TTB. As mentioned above, linear models also rely on assumptions about cue usage. These problems are, however, easier to handle due to the compensatory nature of linear models. For example, not using the most important cue has potentially much greater effect for TTB than for a linear model.

Human Decision Making

Decisions made by humans offer no way to objectively review the decision process. The only information offered for outside parties is the motive the decision maker presents. Not only is this a problem when trying to analyze the decision process, it also poses a problem for the decision maker. In order to learn from past decisions he has to rely on introspection aided almost solely by the arguments he presented to the public and vice versa. This means that the feedback available is less structured than in a mathematical model. Also, it is harder to get a reference point for what kind of process would have given better accuracy, i.e., the decision process of the leading expert in the field. In other words, there is no easy way to compare ones own decision process to a process that would have made a better decision.

As one might assume from the above argument, humans hardly learn at all from our own experience when we make expert judgments without proper feedback. This has been the subject of a lot of research. Most of it has been performed in the field of medicine, and indicates that if we learn from experience, it is only a minimal improvement (Brehmer, 1980; Camerer & Johnson, 1991). This means that if designing a decision aid is too difficult, if not impossible, one could still passively improve decisions by offering structured feedback. This could be done by giving the weights that a linear model would have used in order to make the same decisions as the decision maker. Even if one does not know the optimal weights for the model, presenting the decision maker's own weights would benefit introspection into the decision process. The benefit in this case, comes from presenting the relative importance of the cues.

Transparency into how the decision process differed from a process leading to a more correct judgment is an important tool for motivating the use of decision aids. People will become used to them and see that also their unaided decisions improve. Using transparency in such a way requires that the model being used to give feedback is easy to understand.

Consistency

Consistency is a major concern in all decisions involving justice. If the same set of circumstances does not lead to the same decision, the decision process will be seen as random. This should be avoided when the decision making body has to be perceived as fair and respectable. Courts, for instance, has to be seen as trustworthy and reliable for people to respect the law.

One major problem when trying to attain consistency is irreducible uncertainty (Hammond, 1996). Despite our best efforts to take in all the facts when making important decisions, those facts will always be tainted by our senses, how they were presented to us, and how we formulate them to fit into our arguments. This means that given the exact same facts, we may make a different decision, regardless of whether we use decision aids. Therefore it is important that the decision maker is aware of this problem and tries to compensate for it. Decision aids can create awareness about uncertainty since they require that the information entered adheres to a strict syntax. This makes the user think twice about what he actually knows. This could backfire though, if the decision maker feels more sure of his knowledge when he has formulated it properly. The syntactic complexity needed to input information into the decision aid is a big obstacle in design. This is one of the reasons for why most decision aids available today have fixed cues. The problem is that most cues are more or less related, and therefore best formulated in probabilities. The most intuitive means of input is natural language. Natural language, however, does not describe probabilities very well, and it is difficult to input relations without using mathematical language. The best means of input may therefore be through graphical structures, e.g., connecting nodes that are related, and describe the relation through the shape of the connection.

When trying to achieve consistency through statistical analysis one reduces the ability to handle rare cases. The reason is that the statistical analysis will extract the standard case from the data. This is an important aspect when it comes to models of decision making; the reason is that they will always give the same advice given the same information. Therefore the reduced ability to handle special cases becomes more important than it is in human decision making, which often lacks in consistency to begin with.

Linear Models

The consistency of linear models depends on the procedure with which the weights are calculated. The most common method is to use multiple regression. Doing so results in the best fit to the available data. The downside to using multiple regression is that very important cues that occur rarely will receive low weights. These situations belong to the “broken leg” (Meehl, 1954) category of problems, indicating that if someone breaks their leg, prediction of their daily behavior as an extrapolation of their normal behavior is not possible (Dawes et al., 1989).

Since one cannot solve the problem of broken leg situations by using linear models, one has to use computer programs that investigate such cues first, as a means to handle the non-linearity. This could be done by simply adding if-statements before the actual application of the model. This would not necessarily lead to decreased performance if broken leg cues were ordered by importance. In principle, however, one cannot have high consistency and still be able to handle broken leg situations in linear models. The reason is that the consistency comes from the calculation of the standard situation; the further one gets from the average, the worse the performance becomes.

It should be noted that one cannot handle all possible broken leg problems, there are simply too many. However, one could handle the most common problems that are not extracted in a regression analysis, without too much effort. A more sophisticated approach would be to have a model that is capable of detecting that none of the advice given fit the current situation.

Bayesian Networks

When building a Bayesian network statistical analysis is used to calculate the proper connections between the nodes. Since these connections are specified for every combination of nodes one can also add nodes to handle broken leg situations. This means that one can achieve the same level of accuracy as other models based on statistical analysis and still handle rare information without any loss of consistency. The method of adding nodes to handle rare information is of course similar to adding if-statements to a program based on linear models. In a Bayesian network, however, this can be done in a probabilistic fashion. This means that the other information available will not be overridden as it will when using simple conditional statements. The inherent problem with this approach, however, is that one has to specify a lot of broken leg nodes in order to handle most of the rare occurrences in the situation.

One of the major advantages of using Bayesian networks is that the information also propagates backwards. Therefore one can identify the cause of the decision situation. For example, if lung cancer is detected in a patient, propagation of this information in the network will give the probabilities of causes for lung cancer. This makes it easier for the user to see what signs may have been overlooked when the cancer was not diagnosed.

Take The Best

Just like linear models, TTB will behave exactly the same give the same set of information. However, when the information differs it is harder to assess what effects it may have for TTB than for linear models. The reason is that there is no guarantee that the new information will be used at all. For instance, if we learn that the third most important cue in fact is irrelevant, the only way that might change the decision is if the advice process actually uses the three most important cues. The probability of it being used can be calculated by comparing the validity of the specific cues, though. However, just as with the predictability in the section on transparency, it is harder to assess the consequence of the difference in information. Not being able to assess what effects difference in information will have is generally not a problem if one can perform simulations at a low cost. Therefore this deficiency only poses a real problem when developing the model, and has to put it through rigorous tests. For instance, when testing what cues to use, one will have to run a lot of simulations, increasing the overall cost of the development.

Since TTB is non-linear by nature it is easier to modify it to accommodate broken leg situations. Doing so, is simply a matter of putting the broken leg cues first, despite their actual ecological validity. This, however, would defeat one of the major advantages of TTB; speed. By looking at only a few cues at a time it accesses less information than any model based on multiple regression would. But as mentioned previously information access and simpler calculations do not necessarily have any bearing on whether the model is better suited for computer execution or not.

Human Decision Making

The consistency of human decisions differs a lot from person to person. However, since we do not have access to the exact reasons for our previous decisions, it is reasonable to assume that the maximum consistency we can reach in expert decisions is less than that of a computer program. This follows from that our memory is not perfect, meaning that we cannot follow an as narrowly defined policy as a computer can. This lack of perfect insight into our previous decisions may, however, be the reason for why we are able to handle problems of the broken leg category much better. We are forced to reconsider not just the previous decision, but also the information on which we based it. This makes it possible to extrapolate a general policy, as well as a local policy. The general policy is the overall policy we follow in a particular category of decisions, and the local policy is the policy we have chosen to follow in the cases most similar to the current decision situation. In other words; not adhering to a strict policy means more consideration of the facts and hand, which in turn results in a better handling of uncertainty and rare information.

In decision aids where humans input the information, it is likely that the decision maker will detect any broken leg situation and compensate for this. Therefore it is possible that the decision aid and its user together will be able to maintain the high performance in uncommon situations that humans have, while increasing the consistency in standard decisions.

Accuracy

The most basic requirement of good decisions is that they are somewhat correct; this is measured by accuracy. When designing simple decision aids, accuracy is by far the most important aspect of the advice given. However, when the consequences of the decision grow, the other aspects become increasingly important. Choosing what kind of cereal to buy, for instance, does not require the same amount of transparency, or accuracy for that matter, as a court ruling.

When measuring accuracy one can either look just at the percentage of correct advice given, or calculate it by weighing the different kinds of outcomes differently. This means that one weighs the different outcomes in the Taylor-Russell diagram differently, prioritizing the reduction of one kind of injustice. For instance, one could give a higher weight to false positives than false negatives in the case of court decisions. Since such weighing is dependent on policy, this will be presented in further detail in the section on adaptability.

The biggest problem in evaluating the accuracy of different models is that one needs to know which advice was right and which was wrong. Most decisions do not offer the possibility of assessing what would have happened if another decision had been made instead. Therefore such studies are usually difficult and take long time. In the field of medicine, predictions of, e.g., life expectancy is one way to see how correct the doctors were. In diagnosing mental illness, however, a healthy patient sent to a mental institution may very well become mentally ill from the environment before one finds out if the doctor’s decision was correct or not (Dawes, 1988). This is why it is easier to create decision aids for situations where there are enough cases of different diagnoses and treatments to calculate which diagnosis is correct. This, however, is complicated by the fact that several diagnoses may result in almost the same treatment. Although this makes it easier to give the correct advice, if the advice is given in the form of treatment and not diagnosis, the actual accuracy will be lower. This especially affects adaptability.

Linear Models

There have been many studies comparing human judgment to linear models; most of them have been done in the field of medicine (Meehl, 1954). What they all find is that, linear models consistently outperform human decisions in accuracy. It turns out that one can basically use any weighing of the cues as long as the weights have the right sign. It is only when the linear model has trained on data from about 15 times as many cases as cues that a weighing based on multiple regression outperforms unit weighing (Dawes & Corrigan, 1974).

Although it may seem strange that random weights render a good result, this result is seen as trivial since there is a strong correlation between random weights and optimal weights (Dawes, 1988). This means that as long as the sign is correct; one will achieve a fairly correct linear model, without doing a lot of work. The reason for the correlation is that cues are rarely independent. Therefore a cue being weighed too low will to some extent be compensated for by too high weights for other cues. The less dependence there is between the cues, the lower the compensating effect will be.

Bayesian Networks

Since Bayesian networks can be built using a similar statistical analysis as linear models they have roughly the same level of accuracy. However, since the structure of the network and the math behind them can be quite complex, there is an increased risk of errors entering the system. This, though, is a human error and cannot be blamed so much on the model as the designer.

The probabilistic approach of Bayesian networks makes it possible to present the user not just with the advice that matched the information best, but also the probability for the other possible advice. This means that one can immediately see if the information supports any other advice. For instance, if five diagnoses are possible and one receives a probability of 0.4 and two others 0.3, we know that the recommended diagnosis is by no means definite. These probabilities also make it easier to see if there are any patterns in the diagnosis, such as two possible diagnoses being much more probable than the others. This clear representation of the closeness between two advice should lead to an increased understanding of the concepts that the advice constitutes. For instance, two seemingly different diseases may give roughly the same symptoms. Knowing this will make the decision maker more aware of the risks of misdiagnosis. Such transparency may improve accuracy since the decision maker will be more cautious.

Take The Best

The accuracy of TTB has usually been measured in simple selection tasks, such as assessing which German city is biggest. Gigerenzer et al. (1999) has tested TTB against other models in 20 such selection tasks. In these experiments it turns out that TTB performs equally well to models of multiple regression. The reason for why TTB can perform as well as multiple regression in simple tasks is that it in essence is a sequentially executing weighed tallying. Since the weighing is done by calculating the ecological validity, "the relative frequency with which the cue correctly predicts the target" (Gigerenzer & Goldstein, 1996), the weights are better than random weighing. This combined with the fact that recognizing the city at all is highly important, means that the possibility for increased accuracy by more advanced models is quite small. In the experiment referenced here, the performance gap between linear models and TTB disappeared when the linear models were allowed to take advantage of this *recognition heuristic* (Gigerenzer et al., 1999). This suggests that accuracy in this experiment is highly dependent on a single cue.

Human Decision Making

As mentioned above; models of decision making, e.g., linear models, almost always outperform human judgment. However, the studies of this have been done in specific areas of decision making. The strength of human decision making may therefore lie, not in high accuracy in specific subjects, but rather in acceptably high accuracy in many different areas.

The reason for why most studies are done in narrow fields is that it is more difficult to study a general accuracy. In narrow studies, there is no need to identify what information or method applies to the situation. The biggest problem for decision aids is not to be more accurate than humans, but to identify what the circumstances are, and select the suitable model. For instance, a decision aid developed to diagnose a certain illness is only useful if the doctor knows, or suspects, that the decision aid is applicable. In order to make decision aids more useful they should be able to handle all kinds of different situations at the same time, and find the most likely diagnosis given the symptoms. This, however, is very difficult, which is why there are essentially no general purpose decision aids today.

Improvement

When designing decision aids, a high accuracy is important, but it is also important to use a model that can improve easily. Without improvement when possible, the performance will never be as good as it can be. If possible, the enhancement process should be automated, requiring no reprogramming or redesign. Therefore it should be possible to improve decision aids with a minimum of effort. One way to make it easier for the decision aid to improve is to make sure that it does not need to store all the old data that was used to create the model in the first place. Storing old information would not only require a lot of memory, it would also pose a risk to security and privacy. For instance, in order to recalculate the weights of a linear model, one does not want to store all the old data. Using the old weights as a starting point would reduce the memory needed, and reduce any loss of privacy for the people who contributed to the original data.

Human decision making has had a long time to develop. All improvement through evolution, however, lacks a clear direction and is quite disorganized. This means that, while the basics for good decision making exist, it has probably not been brought to its full potential. Scientific attempts at improving the human decision process have only taken place quite recently. One such attempt is the development of decision trees for detecting the risk of heart attack by Breiman et al. (1993). It also seems that extensive training in statistics result only in a slight improvement in assessing probabilities in everyday situations (Kahneman et al., 1982; Gilovich et al., 2002). This means that we are not just bad at learning from experience; we also have trouble learning new ways of thinking. This is something that decision aids might be able to improve, even if the users are not willing to base their decisions on the recommendations. The reason is that the user will have to motivate not just his decision but also his disregard of the advice given to him. Such a motivation does not have to take place in the decision aid, but rather it will take place on an internal level.

One aspect of improvement is the initial training of the model. The fewer cases needed to improve the model, the easier it will be to construct a good decision aid. The more difficult it is to assess other possible decisions, the more important it is to only have to look at, and evaluate few such situations when constructing a decision aid.

Linear Models

Linear models are difficult to improve. As mentioned previously, linear models with weights determined using multiple regression do not really differ in performance from models with random weights. Only after about 15 times as many cases as cues does the difference in accuracy between a unit weight model and multiple regression model become significant (Dawes & Corrigan, 1974). This means that while linear models are capable of attaining a high accuracy, they need a lot of data to train on. This becomes increasingly problematic the rarer a decision is, which is when decision aids can improve decision making the most.

Although linear models require a lot of data to reach good accuracy, it is not necessary to store old data in order to improve when new information becomes available. One can simply use the old weights as a starting point for calculating new ones. This also means that the process of improving the performance requires fairly few calculations.

Bayesian Networks

Bayesian networks require only a few cases to stabilize. The reason is that there is a structured approach to the statistical analysis. Linear models, on the other hand, use a brute force approach, i.e., try every possible relation. The network structure is not only an advantage for the improvement process, it also makes it easier to test the model, and make sure that it is correct.

Just as with linear models, there is no need to store old data in order to retrain the network when new information becomes available. There is also no need to generate a new junction tree for the network. This is especially important since finding a junction tree is potentially NP-hard.

Take The Best

The learning curve for TTB is about the same as for linear models. TTB, just as linear models, needs more samples to train on than Bayesian networks. Unlike linear models, cues cannot be randomly ordered. Having incorrectly ordered cues in TTB is potentially much worse than having improperly weighed cues in a linear model. There have, however, been some experiments with random cue ordering, that shows it may not be devastating (Gigerenzer at al., 1999).

Since TTB relies on relative frequency of accurate prediction it does not need to store old data. The only values that need to be stored are cue validities, which can be recalculated when new information arrives.

Human Decision Making

Most studies on human judgment show that in expert decision making, we hardly learn from experience (Brehmer, 1980). However, given proper feedback, improvements in human decision making does occur (Wigton et al., 1986). This means that the lack of improvement lies in identifying what information is relevant rather than in some generic inability.

Even though the human memory does not seem to have the same kind of restrictions as computers, it has some drawbacks. Humans have a tendency to give a higher relative weight to rare occurrences than to those that happen all the time. For instance medical doctors remembered the times when a linear model was clearly incorrect better than all the other times it was not (Dawes, 1994). This means that the problem is not one of capacity, or of recalculation, but rather of making sure that the information that we analyze reflect reality. Something that can be aided by the use of decision aids for more detailed feedback.

Adaptability

Since the four different outcomes (in a Taylor-Russell diagram) of a decision are linked to each other, and decreasing only one kind of error inevitably increases the other, it is up to the people setting the current policy to decide how the outcomes should be weighed. The two different kinds of errors are false positives and false negatives. In the case of a court decision, a false positive is when an innocent person is sentenced, and a false negative when a guilty person is freed. The first kind of error poses an injustice to individuals, and the latter poses an injustice to society. Therefore, when society "gets tough on crime" this means that the cutoff in the Taylor-Russell diagram is moved to the right, meaning more innocent people will be sent to jail. In the example of hospitals, if a certain diagnosis costs the hospital a lot more than a clean bill of health, there is an economic incentive not to diagnose the illness. If the illness is a serious one, the hospital economy collides with the patient's wishes. In such situations the hospital's policy is a determining factor for the decision. This means that the decision aid has to provide a comprehensible way to weigh the outcomes differently.

Although it may seem questionable to perform outcome weighing for purely economic reasons, the errors that are being made today go both ways and the only way to reduce both kinds of injustice is to improve the accuracy of the decisions. Therefore using a good decision aid will reduce the injustice to all parties involved, making the weighing have less consequence, no matter which side you prioritize.

Linear Models

The simplest way to make linear models weigh outcomes differently is to change the multiple regression. This can be done by adding weights to the outcomes, and then let the regression use these weights when measuring accuracy. The decision aid can be designed to make the addition of weights intuitive by allowing the people who set the policy for the specific decision aid to use graphic tools to identify and set the weights they feel are appropriate. Simplicity is especially important when the policy setting may be especially controversial. By using computer programs that show the explicit outcome changes that occur with different weightings one can let the decision makers do this explicitly, without involving engineers or other personnel. Letting the policy setters themselves decide the consequences is important for accountability, which should lead to more careful consideration of the consequences of the decisions.

Bayesian Networks

In a Bayesian network it is even simpler to adapt advice to policy. The reason is that the probabilities for all different combinations can be adjusted in the advice node. However, given the more complex math behind the propagation between other nodes in the network, it may be hard to adjust for different outcomes without loosing accuracy or coherence. Therefore it may be better to simply let the network train itself, using different weights for the outcomes, in a similar fashion to the weighing in the multiple regression in linear models.

Take The Best

Since TTB relies on frequencies of correct decisions, it is quite easy to modify it to match the set policy. Just like linear models, the procedure is straightforward enough for novices to set the policy using simple graphical tools. In contrast to linear models, however, the result will be a reordering of cues, rather than a different weighing. Just like the use of random ordering, described in the section on improvement, this leads potentially to a decreased accuracy. Accuracy in this case refers to the models adherence to the definition of accuracy set by the policy. That is, if we weigh false negatives higher than false positives, TTB is more likely than linear models to loose accuracy in the two correct cases. In such a case, the reduction in accuracy depends not on the weighing of the outcomes, but on the average cue validity during the execution of TTB. That is, weighing the outcomes differently will lead to an overall decreased cue validity, resulting in a lower accuracy or lower speed.

Human Decision Making

The major advantage of human judgment, with respect to adaptability, is that we understand the policy behind different weighing of outcomes. However, we are often affected by the consequences of the decisions we make. It is therefore likely that while we may believe in a certain policy, we stray from it when the people affected by the decisions are close to us. Therefore we may be good at adapting our decisions to policy, but at the price of consistency.

Another problem with human involvement in the decisions is that different agendas get more room. This means that the central policy set by the people who are ultimately responsible for the consequences may not be obeyed by the people making the decisions on the lower levels. Since free thinking and ambition are not necessarily bad it may be a bad idea to force people to follow the advice of a decision aid. However, by making people motivate their disregard of the advice one can achieve a more coherent policy without sacrificing the versatility that differentiating policy brings to an organization.

Speed

There are a lot of models (Juslin & Persson, 2002; Gigerenzer et al., 1999) that premier speed instead of accuracy and consistency, TTB is such a model. Although speed undoubtedly is an important factor in many everyday situations, such as swerving to avoid hitting an animal on the road, it is, compared to accuracy, not that important for expert judgments. Even though fast decisions in a hospital emergency room is important, a lot of medical decisions require high accuracy rather than high speed.

An evolutionary perspective is often used to motivate the need for speed in decision making. Most people alive today presumably had ancestors who made fast and somewhat accurate decisions. However, a lot of historically crucial decisions, such as when to plant crops, do not become bad decisions if they take a few hours instead of a few seconds to make. This means that taking all the time one has before the consequences of the delay become too great would be the ideal decision making speed, just as using all the material available to improve something is good engineering practice. Since it is almost impossible to assess the effect of slower decisions, a good margin of time seems preferable. That is, one wants to use as much time as one has, but one cannot properly assess how much time that is. Using a fixed time procedure to get a rough consequence estimate, and thereby the time available, is necessary for this. Otherwise, one would end up in an infinite regression of estimations.

When designing decision aids, one of course has to make them as fast as necessary, but trying to limit calculations because they would be too complex for the human decision making process is not necessary. This means that while one uses bounded rationality (Simon, 1955) to some extent, the bounds are quite different from those of the human mind. At some point, however, the complexity of the advice process will reduce transparency.

Measuring speed in computer models is difficult. Not only is the performance hardware dependent, but there is no single measurement that tells the whole story. The factors that indicate speed are: time complexity of the algorithm, memory accesses, and clock cycles used on a specific computer.

The time complexity is a measurement of how the time requirement increases as a function of the number of data elements to be processed. It is most commonly given as the worst case development in *big O notation*. In the case of O(n) the time complexity is linear, in the case of $O(n^2)$, it is quadratic. For example, traversing a list of data elements has the time complexity of O(n), where n is the number of elements in the list. Of specific interest is if the algorithm is NP- complete or NP- hard, i.e., if the time complexity grows exponentially (<u>N</u>ot <u>P</u>olynomial).

The problems with measuring the number of memory accesses are that computers today load more data from memory than is actually requested by the program. The choice of what data to load is determined by prediction algorithms in the hardware. There are also several layers of cache memory that operate at much higher speeds making the actual number of data blocks accessed a meaningless measurement.

Looking at the number of clock cycles used in the advice process removes the danger of overestimating the complexity of the calculations being performed. However, one instead gets the problem of implementation quality as well as problems of optimal algorithms. Besides, if one actually had the decision aids implemented, one could simply measure the time used by the advice process.

The time complexity of the algorithm is a better measurement than the other options since it is platform independent. There are, however, two complications: optimality and heuristics. Only if one mathematically can prove the algorithms used by a model are optimal with respect to time complexity, can one truly compare the performance of different models. The reason is that otherwise there is no guarantee that the algorithms being compared are both the best possible. If they were not it might be like comparing apples and oranges. This relates to that one can often forfeit the guarantee of a correct answer and use an algorithm based on heuristics to get a lower time complexity.

In conclusion; speed in computer models will be compared as a rough assessment based mainly on memory usage and algorithm complexity.

Linear Models

Linear models, where the weights are calculated through multiple regression, require quite a large amount of memory accesses to perform calculations. Also, during the advice process, there is extensive data access. Specifically memory usage is constant; even though it may have been enough to look at only the most important cue, all the cues are weighed together for all advice. The time complexity of multiple regression is quite acceptable, and so is the time complexity of the weighing. Specifically neither algorithm is NP-hard, i.e., exponentially increasing with respect to the amount of data.

Bayesian Networks

Although most of the algorithms for Bayesian networks are NP-hard, one can use heuristics to minimize the negative effects of this. However, there is undeniably a higher risk of getting a very long response time than the other models. The reason is that there are many steps in the algorithms that are NP-hard (Cooper, 1990), whereas most of the work done in linear models and TTB can be done before the actual advice process.

Bayesian networks perform many advanced calculations, but most of the variables are the same. Therefore the algorithms benefit heavily from effective caching and does not need to perform as many memory accesses as otherwise required.

Take The Best

In contrast to linear models and Bayesian networks, TTB only uses the cues that are necessary to give an advice. Therefore the algorithm will on average need less information and memory accesses to finish. Before the advice process, however, calculations of cue validities are performed. These calculations are similar to the multiple regression step in linear models, and are similarly heavy in execution. Therefore the benefit of looking at fewer cues is reduced when looking at the whole picture. There have been attempts to create more efficient models than TTB. PROBEX is one such model (Juslin & Persson, 2002).

An important advantage to TTB is that one only needs to input as much information as is necessary for the advice process. This means that if the gathering of information takes longer than the calculations of the decision aid, the non-compensatory nature of TTB is an advantage. Bayesian networks have a similar advantage in that the probabilities of the different advice are updated continuously. This gives the decision maker the option of terminating the information gathering as soon as an advice seems likely to win.

Human Decision Making

In everyday life humans make decisions very quickly, e.g., choosing what cereal to buy is simpler to us since we know that the consequences of a bad decision are minimal. Making a similar, or even the same, decision often means we do not have to think as hard.

Understanding the consequences of a decision means that one can assess how much time is reasonable to spend on making the decision. One way to make models operate faster is to make them adapt to the importance of the decision. For instance, if there are few consequences, and therefore lower accuracy requirements, the decision aid can prioritize speed. However, in decision aids the input phase of the advice process usually requires so much time that speed is secondary to accuracy. Therefore such an understanding is pointless unless the information that the advice is based on is gathered automatically, i.e., in an artificial intelligence.

The main difference between experts and laymen is not that experts look at more information to make their decisions; it is that they seem to have better structured memory (Camerer & Johnson, 1991). This reorganizing of memory to better suit the decision process is quite similar to how TTB and linear models work. However, there is currently no way to know if the information is actually stored differently in the brain or if the heuristics in the brain change.

Summary

Summarizing the performance of the models and human decision making with regard to the different aspects is best done in table form. However, determining which model is the best is mainly a subject for the discussion.

Aspect	Linear Models	Bayesian networks	Take The Best	Human Decision Making
Transparency	Excellent	Very Good	Excellent	Fair
Consistency	Excellent	Excellent	Excellent	Fair
Accuracy	Excellent	Excellent	Excellent	Good
Improvement	Good	Excellent	Good	Fair
Adaptability	Very Good	Excellent	Very Good	Good
Speed	Good	Fair	Good	Very Good

Table 1: Summary of the performance in the different aspects.

It should be noted that one cannot fully summarize the performance into just one term; instead it is the relative performance that is of importance. Note that throughout this thesis we have focused on decisions requiring some level of expertise. In everyday decisions human decision making would probably be more competitive. It is also noteworthy that the transparency of human decision making depends on the decision situation and the people involved in, and affected by, the decision. It therefore ranges between very bad and good, averaging approximately on fair.

Relations between Aspects

This is the most important topic of discussion, because there is not really any experimental way to find out how aspects should be weighed. The main reason is that it requires quite a lot of introspection to identify how one personally would weigh these aspects. The weighing also depends on the situation in question. In the following sections, the aspects have been put under different headlines to improve the structure. Since relations always include at least two aspects, this sorting is to be taken mainly as a guide.

Transparency

Transparency is related to improvement in that it is easier to improve future decisions if there is proper transparency. Without a description of the decision process, there is no way to duplicate it. Lack of insight into previous decisions essentially means that future decision makers have to start from scratch.

Speed and transparency are related since the more information about the decision process has to be stored, the slower the decision process will be. The slowdown appears no matter how one measures speed. For instance, the number of memory accesses will increase, as will the actual execution time.

Consistency

Consistency is related to transparency in that a person that understands the decisions will be able to more correctly assess if they are consistent or not. The reverse also applies; an increased consistency makes it easier to understand how the decision was made.

Consistency is also related to adaptability and improvement. The relation to adaptability is that when policy changes, the new decisions will be inconsistent with previous decisions. Also, as the accuracy improves, inconsistencies will occur. Both these relations mean that consistency needs a “configurable start”. This means that when policy changes, one resets the previous consistency for that specific field the policy change took place in.

Accuracy and Speed

Accuracy is related to improvement and adaptability. These relations, however, will be dealt with under those sections. Accuracy is also related to consistency; if there is a high level of accuracy over time it is more likely that the same decisions will be made again, increasing the consistency of the decisions.

One of the main dependencies, or colliding goals, is between accuracy and speed. It is generally assumed that the more time it takes, the more accurate decision one will make. However, this is not necessarily true; not only is it difficult to define what “fast” actually means, but it is also often difficult to define accuracy. The first problem is related to the definition of the decision process. If one includes the time it takes to perceive, identify, and formulate the information then this is probably at least as time consuming as making the decision after the information has been formulated. This means that there is a lot of time to be saved by starting the decision process as soon as information begins to become available. This stands in contrast to waiting until all the information has been made available. TTB is an example of the former, linear models is an example of the latter. The problem of defining accuracy lies, to a great extent, in that we often will not know what would have happened if we had acted differently. Therefore, we cannot define the accuracy of the decisions we did not make. Most likely this is the reason for why people are generally quite conservative when it comes to trying new methods or products if there is “nothing wrong with old one”. One way that this conservatism affects speed is that the more we strive to make the same decisions the faster we will be able to make them. The more experience we have with a situation, the better we get at identifying what could set the situation apart from previous ones. If the situation turns out to be almost exactly the same as a previous one, we can simply make the same decision again, provided that the old decision was not incorrect. This effect can be seen in that most experts improve their speed rather than their accuracy, the more experienced they get (Camerer & Johnson, 1991).

Improvement

Improvement is related to transparency, in that it easier to improve the more transparency there is. Without proper insight into our own, and other’s, decisions we cannot improve as much as possible. We also need accuracy in order to improve, the accuracy in question, however, is a form of meta accuracy, i.e., accuracy about accuracy. This is related to the discussion above about how to define accuracy. If we cannot tell whether the accuracy of a certain behavior is higher than the accuracy of the current behavior, we will not improve at the highest rate possible.

Adaptability

Accuracy and adaptability are intrinsically linked. It is the policy of the situation that defines what we mean by accuracy. Therefore a higher the level of adaptability means a higher level of potential accuracy. However, the higher the adaptability, the faster policy can change, meaning that fewer of the previous decisions may have followed the same policy. In that case one will not be able to improve as much from studying previous decisions. This means that if policy is allowed to change too fast, accuracy will most likely decrease over time. Such a changing of policy also affects transparency; if the policy changes are drastic, the understanding of previous decisions will decrease, leading to a reduction in perceived transparency.

The Models from a Search Tree Perspective

Search tree representation is important since it provides a simple way of understanding the models from a computer science perspective. It also makes it easier to understand how the models actually operate, and how this affects the performance in naive implementations. Often there are more advanced data structures that make more complex, but faster, algorithms possible. One example would be a data structure that transforms the information into mathematically advanced sets and operates on these. The junction tree algorithm for Bayesian networks is a good example of this (Jensen, 2001). However, such algorithms may be too difficult to understand and too expensive to implement for most situations. This naturally depends on how many different situations the decision aid is intended to handle. The more situations, the bigger the market, and the more money will be available for development. But since there are few professional decision aids available today, development will probably begin at a smaller scale.

When building a decision aid one has to design it either to be applicable in many situations, or as an aid to experts who already have a fairly clear idea of what they are looking for. For instance, a decision aid capable of giving several diagnoses would be quite similar to a large search tree. A decision aid, aimed at diagnosing only a few, closely related diseases on the other hand, is more similar to the second last node of such a search tree. Depending on the size of the tree, one has very different efficiency requirements – in a small search tree an algorithm with exponential time complexity is not necessarily a problem.

Ideally one would want a general decision aid that would present an intermediate result, i.e., showing which advice is most probable, given the information that has been acquired. In order to do so, one would either need a lot of precalculated probabilities, or let the probabilities be calculated, and recalculated, as new information arrives. The latter is a more efficient approach if one uses winnowing to prune the search tree. Such an approach would require less memory and less computation. The advantage of using precalculations is that it can be done on a faster processor than what is available in a handheld device. Presenting intermediate results are important since the gathering of information always has a cost, be it in time or money. This means that the earlier one can tell which advice will be given, the faster the decision aid will be. In a sense, this means that the need for calculations is often secondary when measuring the speed. For example, if a patient complains about stomach aches, we can rule out a lot of possible diseases. This can be used to identify which information should be retrieved in order to rule out as many diseases as possible in the next step of the information gathering process. This is somewhat similar to how TTB operates, but as we will see when we look at the models from a search tree perspective, TTB runs into some difficulty when there are several choices.

Linear Models

Linear models do not provide a good way of making the advice process serial. That is, one cannot tell the probabilities of the different advice while information is being acquired.

Therefore we cannot calculate the probabilities of different advice before we have enough information to know that one specific advice will win. In order to improve efficiency, one needs to input enough information so that a compensatory winnowing can take place. Designing such a winnowing process becomes more difficult the bigger the search tree is. This becomes apparent in a graphical representation of a linear model.

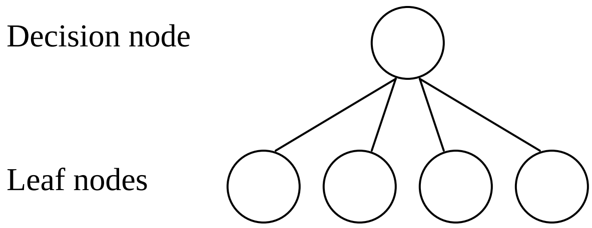

Figure 4: Search tree representation of a linear model.

As can be seen in Figure 4, there is only one decision node, which contains all the information available. The leaf nodes represent the possible advice the linear model can give. All information is given to the decision node, and calculations are performed in the decision node and the most suitable leaf node is selected.

An example of the situation in Figure 4 is a decision aid capable of diagnosing four different diseases. As information becomes available, through tests and other means, it is used as input into the decision node. When enough information has been added for the linear model to be able to identify which disease will “win” a leaf node is selected. This node, containing diagnosis and treatment, is then presented as the advice given to the user.

Since information usually costs both time and money, it is an advantage to not have to gather more information than necessary. One way to make linear model better at giving advice when only a few cues are available is to create a serialized version. Since a tree is a recursive structure, this kind of serialization could be done by chaining several linear models together, letting the layers more than one step above the leaf nodes be decision nodes on how to proceed. See Figure 5 for an example of how to do this for the situation in Figure 4.

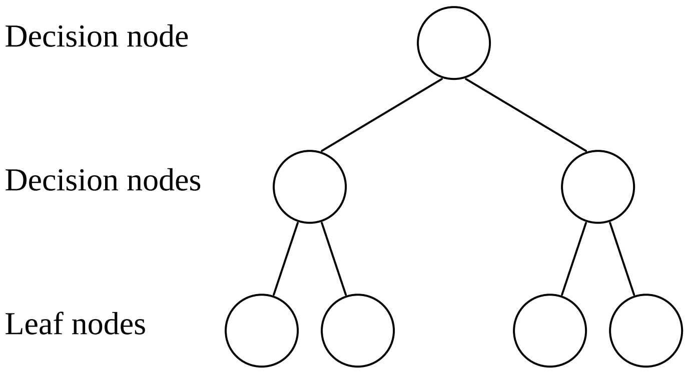

Figure 5: Serialized version of the model in Figure 4.

In the model in Figure 5, one would first input information into the root node, and then decide which way to proceed given the information available. Since we reduce the number of child nodes, we will need less information in each node to decide which way to proceed. Redesigning a tree in this way, however, is just as difficult as finding a good winnowing process. As will be seen serializing linear models is this manner makes them quite similar to TTB.

Bayesian Networks

It is more efficient to not use precalculations if one can perform winnowing based on the probabilities that the current information leads to. Bayesian networks is the most prominent example of such an approach. However, as mentioned in the section on speed, the propagation in the network is an NP-hard problem.

Although a search tree version of a Bayesian network looks just like the linear model in Figure 4, the probabilities of the different advice is continually updated when information arrives. This makes it possible to identify how to proceed with information gathering. Based on the information acquired and the probabilities one can calculate the optimal way to continue the information gathering process.

No example of a search tree generated from a Bayesian network has been presented. The reason is that it is highly context dependent; both the structure and the information available will affect such a generating procedure. It does, however, seem plausible that one can use a Bayesian network as a guide when generating a search tree. By doing so one could generate the optimal tree that can be constructed given the current information available. Since one can auto generate Bayesian networks directly from examples (Cooper & Herskovits, 1992) this makes it possible to create a self optimizing search tree.

Take The Best

Representing TTB as a search tree is somewhat different from the other models since TTB model is inherently non linear. In the simple case of only two possible leaf nodes, the tree looks as follows.

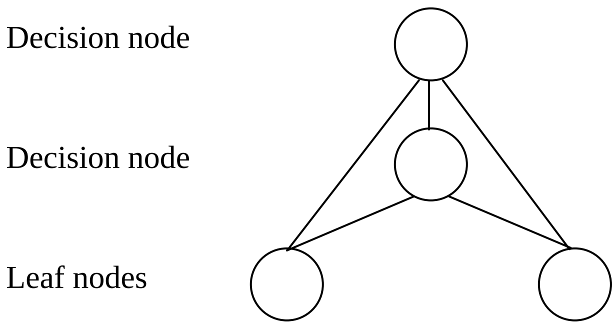

Figure 6: Search tree representation of TTB.

In a diagnosis situation; the leaf nodes are the two possible advice, and the decision nodes the cues available. When receiving information about the first cue, TTB will select one of the leaf nodes if the cue is available for that node and not for the other. If no such difference in information is available, the second decision node will be selected, where the procedure is then repeated. In the situation in Figure 6, TTB will simply guess, if the second node does not differentiate the two leaf nodes.

The more closely related leaf nodes there are, the harder it will be to get a good ecological validity. For instance, correctly differentiating two diseases with similar symptoms is easier than differentiating four equally similar diseases. This is not only because there are more choices, but also because the advice will have more cues in common. In order to handle such situations one needs to chain trees together. Although one can calculate the probability that a certain advice will be given in each of these trees, it will require a lot of information about past decisions. Otherwise one will only calculate the relative frequency of the cues that lead to the different advice. This is a fairly bad measurement since TTB is designed to use as few cues as possible.

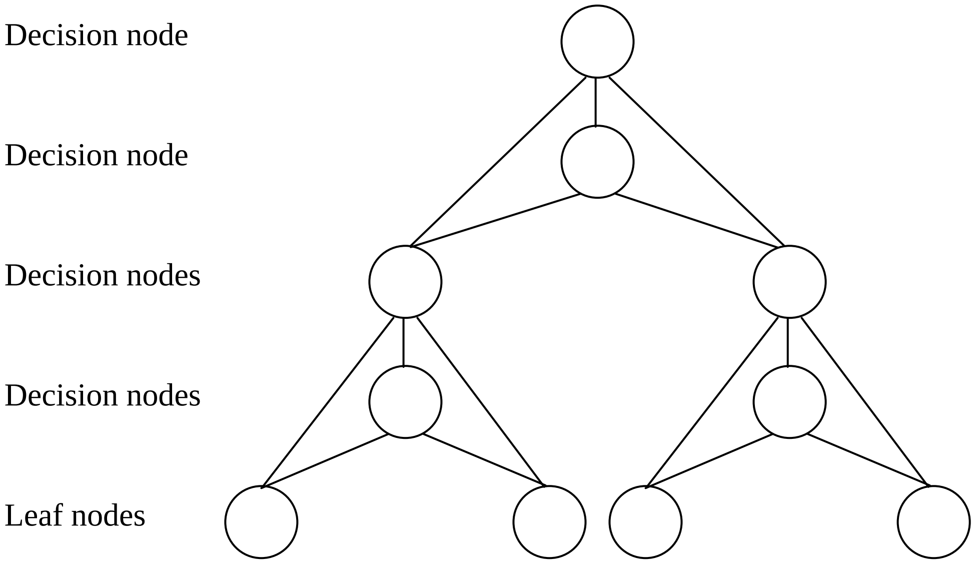

Figure 7: Search tree representation of TTB with four leaf nodes.

As seen in Figure 7, a search tree based on TTB will be very large and contain several duplicate nodes. Another inherent problem is that TTB is non-compensatory, something which becomes more problematic the more leaf nodes there are. The problem is that the more nodes there are, the more nodes will be eliminated even though they may have won in the end. This reduces the possibility of using frequency to predict which advice will be given. This means that the tree will not only be bigger than for a linear model, but it will also be non-compensatory. Furthermore, the calculation of the probable leaf nodes can be performed just as good in a tree based on a linear model. It should be noted that some situations are non-compensatory in nature; in these situations TTB obviously has an advantage.

Summary

As can be seen, serializing a linear model renders a behavior very similar to TTB. There is, however, a major difference, namely that in the linear model the decision will never be based on one single cue. That is, the linear model will proceed all the way through the tree until a separation between two leaf nodes is **inevitable**. TTB, on the other hand, will choose a leaf node when this is **possible**. This means that TTB will still be faster, but at the cost of being non-compensatory. As mentioned before, this cost increases the more advice are possible, unless the situation is non-compensatory in itself.

Examples of Decision Aids

There are a lot of decision aids available for free on the internet. Most of them are based on linear models, but Bayesian networks are becoming more popular. Since there are no decision aids that use TTB, fast and frugal decision trees will be used as the example of decision aids based on bounded rationality. The connection is that fast and frugal decision trees are generated by using a heuristics similar to TTB. The reason why there are no decision aids based on TTB is that it is designed to be a model of the human decision process, whereas fast and frugal trees were designed from the beginning as a decision aid.

Sperling's Best Places

Sperling's Best Places (http://www.bestplaces.net) provides a decision aid for selecting places in the US that would best suit ones preferences about weather, economy, crime, and other factors. It is based on a linear model, and the user provides weights about these factors and then the decision aid weighs the information available on different cities by those weights.

Figure 8: Starting page of the weighing procedure on Sperling's Best Places.

As can be seen in Figure 8, the factor *Weather* has been reduced to a few examples. The internal weighing procedure therefore operates either with several dimensions that the user does not have to weigh, or the weather factor has been reduced to a one dimensional index. One can click calculate at any point during the weighing to make the decision aid calculate the best alternatives using the weights one has assigned and default values for the other factors.

Restart, or modify preferences — Page 1 2 3 4 5 6 7 8 9 10 11

Restart — Modify... — Scoring explanation

(click on city to view profile) — Category Scores

		Climate	Econ.	Hous.	Educ.	Health	Crime	Recr.	Arts	Trans.
1	Seattle-Bellevue-Everett, WA	13%	16%	3%	21%	13%	4%	24%	0%	5%
2	Syracuse, NY	7%	11%	24%	9%	22%	9%	10%	0%	9%
3	State College, PA	6%	16%	15%	17%	17%	16%	1%	0%	12%
4	Bryan-College Station, TX	5%	23%	20%	14%	12%	5%	7%	0%	14%
5	Danbury, CT	4%	15%	7%	15%	10%	18%	23%	0%	6%
6	Duluth-Superior, MN-WI	6%	12%	23%	8%	10%	10%	19%	0%	11%
7	Nashua, NH	7%	20%	6%	15%	19%	18%	10%	0%	6%
8	Rochester, MN	3%	22%	9%	18%	8%	12%	13%	0%	16%
9	Boston, MA-NH-ME	7%	14%	2%	18%	21%	9%	24%	0%	6%
10	Pittsburgh, PA	8%	11%	22%	11%	18%	12%	11%	0%	6%
		Climate	Econ.	Hous.	Educ.	Health	Crime	Recr.	Arts	Trans.
11	Portland, ME	5%	16%	7%	13%	25%	12%	12%	0%	9%
12	Boulder-Longmont, CO	1%	18%	4%	23%	25%	8%	12%	0%	8%
13	Scranton-Wilkes-Barre-Hazleton, PA	7%	11%	20%	4%	26%	19%	3%	0%	11%
14	Naples, FL	3%	21%	3%	31%	7%	5%	19%	0%	11%
15	Charlottesville, VA	7%	26%	4%	19%	23%	8%	4%	0%	9%
16	San Francisco, CA	8%	14%	0%	24%	16%	5%	29%	0%	5%

Figure 9: The result list using the template *Single* on Sperling's Best Places.

In Figure 9 the end result of the weighing procedure is presented. Under most of the headlines in the table there are several sub aspects that one has to weigh. For instance, on the page for *Crime* one can weigh violent crime and property crime differently.

Sites like Sperling's Best Places are fairly common on the internet. They are all limited to one specific field like choosing what car to buy, or what place to move to. The reason is the same for decision aids based on linear models; it is more difficult to create an accurate model the more possible advice there are, and the more different they are.

GeNIe

GeNIe (Graphical Network Interface) is a user interface for SMILE (Structural Modeling, Inference, and Learning Engine). They have both been developed at the Decision Systems Laboratory at University of Pittsburg. Together they form a meta-decision aid, where one can construct, for instance, Bayesian networks. This is done by adding nodes and connections in a graphical fashion. When the network has been constructed one can make inferences and perform tests by changing the information in the network, and set the results that different combinations of information should produce.

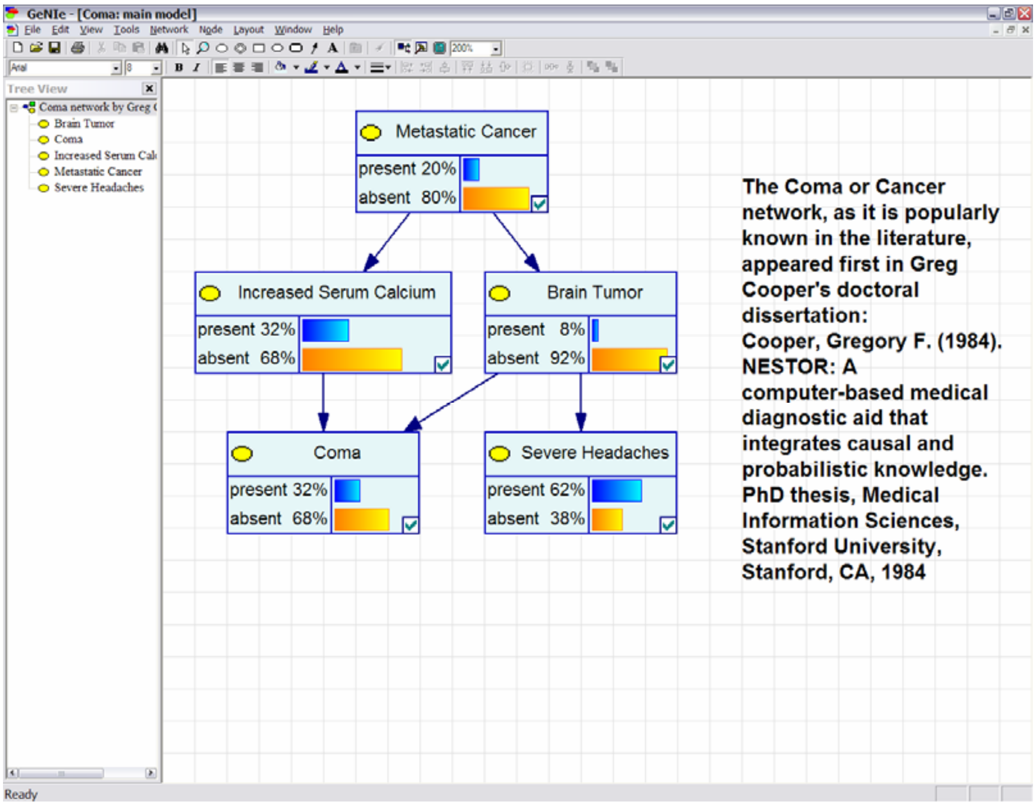

Figure 10: GeNIe with the example network *coma.xdsl* open.

By double clicking on the options, i.e., the colored bars in the nodes, one sets that option to hold true. Then, one lets the information propagate by selecting *Update beliefs* in the *Network* menu.

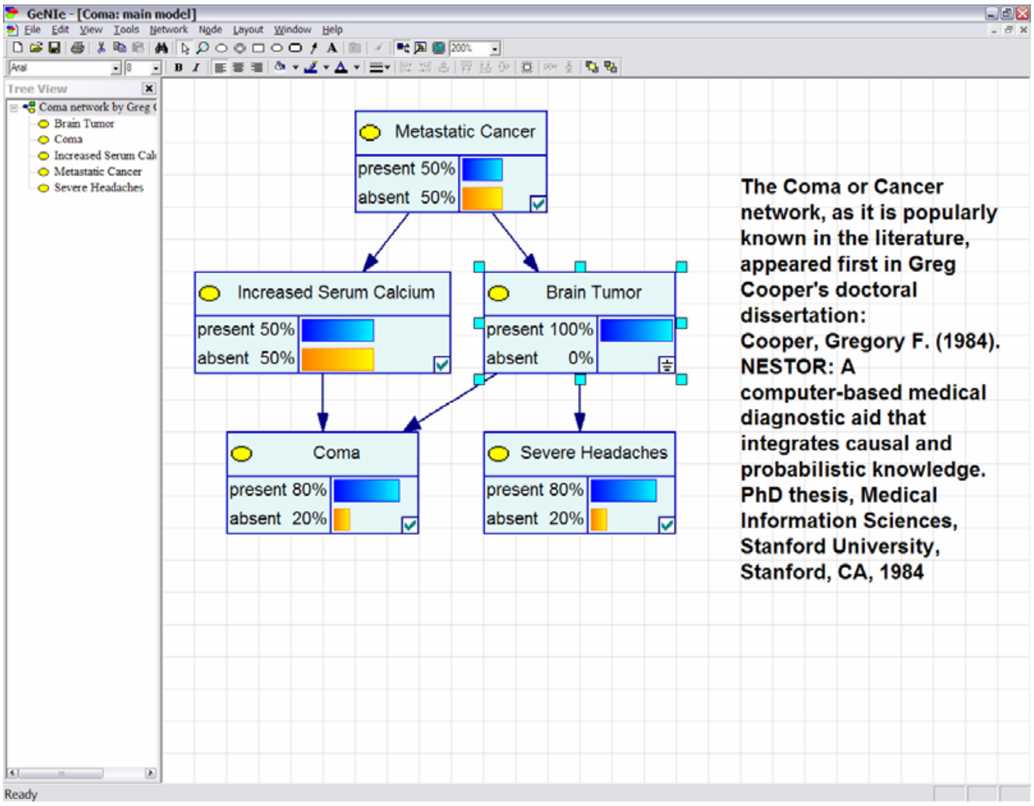

Figure 11: GeNIe with the example network *coma.xdsl* open and added information.

In Figure 11 the option *Brain Tumor* has been set to true. As can be seen the information has propagated to all the nodes in the network. Only by adding this one piece of information can we tell that the probability of coma is very high. Presenting intermediate results in this fashion is invaluable when dealing with a large number of possible diseases and tests to make.

Since GeNIe is a meta-decision aid one can design any kind of Bayesian network and input the information available, and relevant to the situation. The only difficulty therefore lies in designing a correct Bayesian network for the situation in question. There are some limitations to how a Bayesian network can be designed. For instance, one cannot have loops in the network. Another difficulty is designing networks where humans are involved. The network in Figure 2 (on page 6), for instance, suggests that it is only the ice that makes Holmes and Watson crash, i.e., they have no control over the outcome. Interpreting the causality of the network becomes more natural when one gets used to the notation.

Fast and Frugal Decision Trees

The reason for why this master thesis focuses on TTB, when there are currently no decision aids based on it, is that it has been studied far more than fast and frugal decision trees. Also, these trees are built with the explicit intention of acting as a **guide** for the human decision process, which is not the main purpose of a decision aid. The main purpose is to **replace** the decision process. Since TTB was designed to model the human decision process, it is an interesting intermediate between a purely statistical model and the human decision process. As has been shown, there are both advantages and downsides to this approach.

A fast and frugal decision tree differs from an ordinary decision tree in that every level has at least one leaf node. This can be represented in the same manner as the search tree representation of TTB presented earlier. The most well known example of a fast and frugal decision tree was constructed by Green and Mehr (1997). It serves as a decision aid for whether a patient should be sent to a cardiac care unit (CCU) or an ordinary nursing bed (NB).

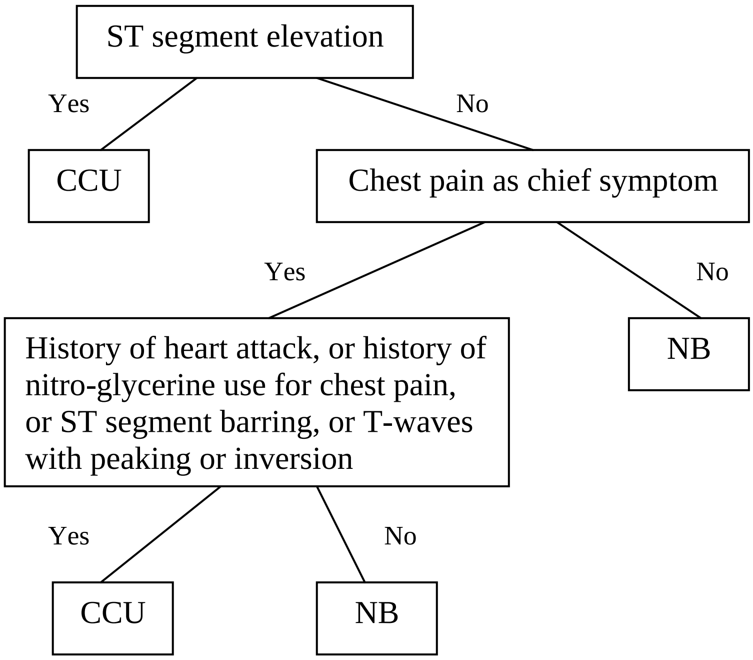

Figure 12: The decision tree developed by Green and Mehr.

As can be seen in Figure 12, there is a leaf node available on every level. Green and Mehr, however, did not clearly state how they came up with the tree. The method for creating fast and frugal decision trees has instead been described in Martignon et al. (2003). The process basically consists of using the cue with the highest maximum of true positives and true negatives.

Creating fast and frugal trees is one of the few applications of heuristics similar to TTB that is available. Also, it is not really intended as a decision aid in the technical sense as much as a decision process for people to follow. This means that one does not take advantage of the computational capacity that a computer-based decision aid can provide. However, the trees are very simple to use and understand. They are therefore a useful “beginners” kind of decision aid. In the article by Green & Mehr (1997), physicians were more likely to use fast and frugal trees than a more complex model based on statistical regression.

Summary

In the three examples above, the performance of the different aspects is as follows.

Aspect	Sperling’s Best Places	GeNIe	Fast and Frugal Decision Trees
Model	*Linear Model*	*Bayesian Network*	*TTB-like*
Transparency	Bad	Very Good	Excellent
Consistency	Excellent	Excellent	Excellent
Accuracy	Indeterminate	Very Good	Good
Improvement	Indeterminate	Very Good	Good
Adaptability	Good	Excellent	Good
Speed	Good	Good	Very Good

Table 2: The performance of the different decision aids with respect to each aspect.

Since the inner workings of the algorithm at Sperling’s Best Places were not available for study, there is no way to know how accurate the decision aid is, or how much training its model needed. It should also be noted that in the case of GeNIe, the performance is to some extent dependent on the network. A larger network than *coma.xdsl*, for instance, may be significantly slower. Since GeNIe is a meta-decision aid its performance is given more generally than simply the performance of a particular network.

Discussion

Since there have not been any experiments in this thesis the evidence instead lies in a detailed discussion. Most of the aspects brought up in this thesis are not very open to experiments. One of the few experiments that come to mind is to thoroughly study human decision making under a long period of time. By doing so one would be able to identify the general performance, even though the variance between decision topics and individuals would surely be quite high.

There are three different topics of discussion that are especially important: how to weigh the different aspects; what model of decision making is best and when; under which circumstances we should use decision aids, and when that would be appropriate. Naturally the four topics are related, but for the sake of structure, however, the main arguments for each topic will be treated in different subsections.

Weighing the Aspects

How aspects are weighed has to be determined by the prevailing policy and the situation in which the decision aid will be applied. This means that finding the best weighing for, e.g., aided governmental decisions require voting by representatives. However, since there are dependencies between different aspects, such an election has to be between different alternatives constructed by experts in decision making. In the situation of a private institution, however, the decisions can be made by, for instance, the board of a hospital.

When the different aspects are weighed, transparency of how the model was designed is required. Without transparency of why certain weights were set, there is not sufficient accountability for the consequences of the aided decisions. Since there will always be some injustice, e.g., sending innocent people to jail, it is important to offer people choices about which injustices they subject themselves to. If there are several alternatives available to, e.g., a patient looking for a doctor, and proper information about the choices, then the patient shares responsibility for any injustices towards him. That is, if the patient could have chosen another hospital with a different policy, but did not, then any involuntary injustices towards the patient were in part brought on by his own behavior. This means that the risk for institutions using decision aids will to some extent be counteracted by good transparency into the decision process.

Since there are so many dependencies between the aspects of a good decision, it is very difficult to identify what would be an especially good weighing. Some trends can be clearly discerned, however.

One such trend is that the more important a decision is, the more need there is for transparency and consistency. Conversely the less important a decision is, the more important speed can become without an unacceptable reduction to accuracy.

Another development is that the more people are affected by a decision, the more need there is for adaptability. Otherwise the democratic order would suffer, leading to a reduction in acceptance for the decision and respect for the decision makers. This, of course, is also related to the importance of the decision.

In economic decisions there is an added need for consistency. If a company is inconsistent in its decision making, the investors will have a harder time assessing future decisions, decreasing the desire to invest in the company. The same also applies to countries on an international scale, where a seemingly random behavior in an important economic matter can lead to a reduced credit rating.

Situations where improvement is important are fairly rare. The reason is that there is less need for a rapid improvement if there is lot of data available. In other words, the more cases one can use to train, e.g., a linear model, the less need there is for a fast learning curve. In situations where there are only a few cases to study, however, one needs the faster learning procedures of, e.g., Bayesian networks. This means that it is mainly in the front line of research, where only a few cases have been studied, that rapid improvement is important.

Since the weighing is so different in different situations there are a few examples below.

Aspect	**Importance**
Transparency	Very High
Consistency	Very High
Accuracy	Very High
Improvement	Moderate
Adaptability	Moderate
Speed	Low

Table 3: The aspect-importance matrix of judicial decisions.

The reason why the need for adaptability is only moderate in the case of judicial decisions is that they are guided to a certain extent by the constitution of a country. The policy affecting judicial decisions are therefore rather static. Also, changes in policy are likely to have an adverse affect on accuracy.

Aspect	Importance
Transparency	Moderate
Consistency	Moderate
Accuracy	Very High
Improvement	Moderate
Adaptability	Low
Speed	Moderate

Table 4: The aspect-importance matrix of diagnosing severe diseases.

There are very different requirements for medical decisions; in an emergency room the need for speed is a lot higher than during regular examinations. In the case of severe, life threatening, diseases accuracy is by far the most important aspect. Also, the adaptability to policy is secondary, since there is often only one course of action. The importance of transparency in this situation could be set to high in order to improve accountability. However, if the accuracy is very high, and adaptability low, the accountability would not be to the patient, but to the hospital. Since the risk of being sued is probably higher if a severe disease was not treated than if it was, the need for accountability is moderate.

Aspect	Importance
Transparency	Very low
Consistency	Very low
Accuracy	Moderate
Improvement	High
Adaptability	Moderate
Speed	High

Table 5: The aspect-importance matrix of choosing what cereal to buy.

For comparison, Table 5 shows the importance of the different aspects in a typical everyday decision. Accuracy is not really that important unless one is very picky about flavor. Improvement, on the other hand, is important since we do not want to make the same mistake twice. Given that most cereal tastes approximately the same, there is little need for consistency and adaptability. Policy on price, however, makes adaptability more important than consistency.

Obviously the need for decision aids is the greatest when the performance of human decision making is lacking in important aspects. Comparing this to the performance of the decision aids (see Table 2) one can see that both GeNIe, and fast and frugal decision trees have a high accuracy. Note also that there is a low transparency in Sperling’s Best Places. However, this is acceptable since there are other available decision aids on places to live (e.g. http://www.findyourspot.com). Naturally the performance of the examples of decision aids are linked to the performance of the models. As we can see, all models fair quite well in the judicial and diagnosing case. In the everyday decision case, however, Bayesian networks is the only one with a fast enough learning curve.

Which Decision Aids are best?

Which decision aid is best is highly dependent on the weighing of the different aspects. It also depends on the situation; whether one wants a general purpose decision aid, or a very accurate decision aid that is only applicable in a specific field.

Linear models, as has been described, have a very high accuracy, high transparency, and high consistency. However, the more possible advice there are, the worse they get, and the more information is necessary. They are therefore suited mainly for specific fields, where the decision maker already has a pretty good idea of what he is looking for. Sperling's Best Places is a typical example of such a decision aid.

Bayesian networks are very fast learners, and have a high accuracy. They are also the only models studied in this thesis that can provide good intermediate results from the information that has been made available. This counters the potential lack of speed that can occur in especially complex networks. Since they can deliver intermediate results, and provide a very clear representation of knowledge they are highly suited for more general purpose decision aids. By calculating the optimal way to proceed in the information gathering process, they provide excellent help in quickly narrowing down the possible advices. GeNIe, being a meta-decision aid, provides all these advantages, and is fairly easy to use and design networks in.

TTB was designed to operate in the way human decision making does. This is to its benefit in situations where human decision making is good, i.e., in everyday situations. Using TTB in expert environments, on the other hand is questionable, mostly because of its non-compensatory nature. However, in situations with few alternative advice it performs on par with linear models. But just like linear models it is doubtful whether TTB is suitable as a more general decision aid. Using heuristics similar to TTB, in the acceptance of imperfections and limitations, is an efficient approach to develop decision trees. Such trees are very easy to understand and use, and are therefore a good way to introduce people to the idea of using decision aids in the first place.

Should we use Decision Aids?

Having established which decision aids are best in what situations, we still need to discuss whether using decision aids is acceptable. In some situations, decision aids may simply be too controversial. This section will explain, and deal with most of these controversies.

As mentioned above, when those who are affected by the decision have made an active choice about which potential injustices to subject themselves to, the responsibility does not lie only with the decision maker. This means that when one can reduce the situation to numbers, and offer people a choice about how these numbers are processed in the decision aid, the ethical problems of using a decision aid are reduced. However, the condition that one has to be able to reduce all elements of the situation to numbers is a controversial issue. One of the biggest markets for decision aids today is in banking. This is because the reduction to numbers is very simple. For instance, when granting loans it is very simple to find a linear model for cues such as how much cash the person applying for the loan has.

The main controversy of reducing people to numbers is that it intuitively seems wrong. However, this kind of reduction takes place all the time when applying for education, loans, or jobs. Therefore the main problem is not the reduction itself but rather that people want "backup options", such as interviews. Research, however, has shown that when using interviews as an extra evaluation procedure, accuracy often drops (Bloom & Brundage, 1947). Because of this, more personable forms of evaluations mainly provide a false sense of justice. This also relates to whether one should aim for a high accuracy on average, or try to make the best decision in every given situation (Bishop, 2000). Since decision aids generally improve accuracy, they actually decrease the injustice to both individuals and society. Therefore, decision aids, despite their reduction to numbers, will actually decrease injustice to everyone affected by the decisions.

Since the only way to reduce both kinds of injustice is by making more accurate decisions, the controversy of numerical reduction lies in making sure that the correlation between numbers and people is as high as possible. For example, one should make sure that the factors involved when granting loans have a high correlation to the actual financial situation of the applicant. One way to increase the correlation is to use several measurements and weigh them together, which is how grade point averages (GPA) are calculated. Often, however, only a few of the grades in the GPA are relevant for a certain college education. Therefore, the GPA should only be calculated on relevant grades. It would also be useful to provide the variance of the grades, since this indicates how focused the students talents are.

Some of the controversy of machine decisions is that they are down to pure and cold mathematics. While this is true, the human brain probably operates in a similar manner, but uses a lot of approximations instead of mathematical formulas. Therefore the argument that machines makes cold hearted decisions really means that it is not the performance of the decisions that matter to people, but rather the effort. This means that people seem to prefer the attempts at making as good decisions as possible, over better and more transparent decisions.

Even though the advantages of using decision aids cannot be ignored, there is still a problem of acceptance. As mentioned previously, there are a lot of decisions where the need for acceptance and respect of the decision maker is important. This means that if the acceptance for using decision aids is low; using them is not an option. Therefore, one needs to gain acceptance in society for the decision aids that one wants to use. The problem is that one has to build the decisions aids first, and show just how good they are, before they can be deployed. This is a complication, since investments with no guarantee of a return require venture capital.

Conclusions

There are three main kinds of models of decision making; linear models, Bayesian networks, and fast and frugal heuristics (TTB). Linear models and TTB are useful in situations with few choices. Bayesian networks, however, are more applicable in situations with many possible choices. The reason is that they provide an intermediate result, offering a way to optimize the information gathering. Although TTB only uses as much information as is needed to give advice, it does not actually provide intermediate results before that point.

Looking at the different aspects that make up good decisions, one can see that in situations of expert decisions, decision aids consistently outperform human decision making. Only when it comes to making non expert decisions, where accuracy is secondary to speed, does human decision making compete with mathematical models.

The aspects of good decisions often depend on, or contradict each other. Therefore they have to be weighed differently in different situations. Finding out which weights are appropriate is a difficult process involving both the decision makers and those affected by the decisions.

There are several reasons for why we should use decision aids. Most arguments against using them do not hold up to scrutiny. One main argument does, however; the lack of acceptance. This means that decision aids can only be deployed in situations where people are not part of the equation, or where a reduction to numbers is commonly accepted.

The general conclusion that has been reached in this thesis is that decision aids actually do outperform human decision making in nearly every aspect of a good decision. They are more transparent, more consistent, and more accurate. They also improve more, and have a higher adaptability. Only in speed does human decision making outperform models of decision making. We should therefore try to increase the acceptance of decision aids in society in general. From that acceptance, the usage will increase, and thereby, the justice to everyone affected by the decisions. And although decision aids may seem impersonal, the reader should by this point have understood that the increased performance lead to a more fair treatment in all but broken leg situations.

References

Bishop, M. A. (2000). In praise of epistemic irresponsibility: how lazy and ignorant can you be?. *Synthese*, 122, 179-208.

Bloom, R. F. & Brundage, E. G. (1947). Predictions of Success in Elementary School for Enlisted Personnel. In D. B. Stuit (Eds.), *Personnel Research and Test Development in the Naval Bureau of Personnel* (pp. 233-261). Princeton, NJ: Princeton University Press.

Brehmer, B. (1980). In One Word: Not From Experience. *Acta Psychologica*, 45, 223-241. Amsterdam, NL: North-Holland Publishing Company.

Brehmer, B. (1988). The Development Of Social Judgment Theory. In B. Brehmer & C.R.B. Joyce (Eds.), *HUMAN JUDGMENT: The SJT View* (pp. 13-40). Amsterdam, NL: Elsevier Science Publishers B. V. (North-Holland).

Breiman, L., Friedman, J. H. & Olshen, R. A. & Stone, C. J. (1993). *Classification and regression trees*. New York, NY: Chapman & Hall.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Camerer, C. F. & Johnson, E. J. (1991). The process-performance paradox in expert judgment: How can experts know so much and predict so badly?. In K. A. Ericsson & J. Smith (Eds.), *Toward A General Theory Of Expertise: Prospects and Limits* (pp. 195-217). Cambridge: Cambridge University Press.

Cooper, G. F. (1990). Computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42, 393-405.

Cooper, G. F. & Herskovits, E. (1992). A Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning*, 9, 309-347.

Dawes, R. M. & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, 81, 95-106.

Dawes, R. M. (1988). *Rational Choice In An Uncertain World*. Orlando, FL: Harcourt Brace & Company.

Dawes, R. M., Faust, D. & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668-1674.

Dawes, R. M. (1994). *House Of Cards: Psychology And Psychotherapy Built On Myth*. New York, NY: The Free Press.

Edwards, W. & Fasolo, B. (2001). Decision Technology. *Annual Review of Psychology*, 52, 581-606.

Gigerenzer, G. & Goldstein, D. (1996). Reasoning the Fast and Frugal Way: Models of Bounded Rationality. *Psychological Review*, 103, 650-669.

Gigerenzer, G., Todd, P. M. & The ABC Research Group (1999). *Simple Heuristics That Make Us Smart*. New York, NY: Oxford University Press.

Gilovich, T., Griffin, D. & Kahneman, D. (Eds.) (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York, NY: Cambridge University Press.

Green, L. & Mehr, D. R. (1997). What alters physicians' decisions to admit to the coronary unit?. *Journal of Family Practice*, 45, 219-226.

Hammond, K. R. (1996). *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice*. New York, NY: Oxford University Press.

Hursch, C. J., Hammond, K. R., & Hursch, J. L. (1964). Some methodological considerations in multiple-cue probability studies. *Psychological Review*, 71, 42-60.

Jensen, F. V. (1996). *An Introduction to Bayesian Networks*. London, UK: University College London Press.

Jensen, F. V. (2001). *Bayesian Networks and Decision Graphs*. New York, NY: Springer-verlag.

Juslin, P. & Persson, M. (2002). PROBabilities from EXemplars (PROBEX): a "lazy" algorithm for probabilistic inference from generic knowledge. *Cognitive Science*, 26, 563-607.

Kahneman, D., Slovic, P. & Tversky, A. (Eds.) (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Martignon, L., Vitouch, O., Takezawa, M. & Forster, M. R. (2003). Naive and yet Enlightened: From Natural Frequencies to Fast and Frugal Decision Trees. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological Perspectives on Reasoning, Judgment and Decision Making* (pp. 189-211). West Sussex, UK: John Wiley & Sons Ltd.

Meehl, P. E. (1954). *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis, MN: University of Minnesota Press.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. New York, NY: Cambridge University Press.

Simon, H. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99-118.

Stigler, J. W. & Hiebert, J. (1999). *The Teaching Gap*. New York, NY: The Free Press.

Taylor, H. C. & Russell, J. T. (1939). The relationship of validity coefficients to the practical applications of tests in selection. *Journal of Applied Psychology*, 23, 565-578.

Wigton, R. S., Patil, K. D. & Hoellerich, V. L. (1986). The Effect of Feedback In Learning Clinical Diagnosis. *Journal of Medical Education*, 61, 816-822.