INVITED PAPERS

Multimodal Interaction : A new focal area for AI

Philip R. Cohen

Center for Human-Computer Communication

Oregon Graduate Institute of Science and Technology

http://www.cse.ogi.edu/CHCC

Abstract*

AI research has often been driven by popular visions – HAL 2001, Asimov's Robot, Star Trek – and by critical application areas – medical expert systems, spoken dialogue systems, etc. These visions and applications serve to inspire and guide researchers, posing challenges, illustrating technical weaknesses, and generally channeling creative energy. Without doubt, the widely held vision of the autonomous robot, has exerted a substantial integrative force, such that numerous disciplines, ranging from mechanical engineering to cognitive science, can see how their intellectual endeavors can contribute to the overall endeavor. In this brief position paper, and in the accompanying talk, I would like to propose that the next generation of intelligent multimodal user interfaces can offer a similar intellectual focus for AI researchers. After providing a brief overview of our work in this area and two examples, I would like to suggest the potential impact that such interfaces could have in the relatively near-term.

Introduction

In this position paper, I would like to illustrate recent work in intelligent user interfaces, specifically the subfield called multimodal interfaces, and argue that multimodal interfaces should occupy a similar place in the AI research endeavor as autonomous robotics. Along the way, I will describe the kinds of skills that were needed to build two multimodal systems in our laboratory, the QuickSet and Rasa systems. As you will see, AI and Computer Science methodologies played a central role, but are informed by cognitive science and ethnographic research and methodologies. This paper is not intended as a survey of the field (cf. [29]), but rather as an example. Numerous other groups, at CMU, DFKI, Illinois, Rutgers, SRI, and elsewhere are also engaged in such research, and the reader is urged to consult their work.

The paper first considers problems of building principled multimodal systems that fuse information at the semantic level. Issues to be addressed include deciding on relevant principles, observing users, and building an appropriate software architecture. Then, I provide two examples, QuickSet, a multimodal pen/voice system for interacting with map-based applications, and Rasa, a tangible multimodal system that enables users to employ paper-based interfaces. In each case, evaluations of the systems by the intended user population have taken place and are summarized. Finally, I describe the kinds of multidisciplinary methodologies that were used in building these systems.

Multimodal Interaction

Multimodal interfaces are those that allow the user to employ a substantial range of human sensory capabilities to obtain and interact with desired information, and to perform tasks. For instance, such interfaces typically enable a user to employ some subset of speech, gaze, body movements, pen strokes, haptics, etc. Multiple modes are advantageous for a variety of reasons, including: robustness, flexibility, ability to correct persistent errors in one mode by using another, ability to avoid expected errors by switching modes, expressivenes, and speed [3, 25, 27, 29]. Such systems are also appropriate to ubiquitous computing environments, small devices, and very large devices.

Principles

Because of the complexity of building multimodal systems, we choose to investigate a principled approach to their design. The principle we have adopted is to use each modality for its strengths and to overcome weaknesses of the other(s) [5]. By studying the strengths and weaknesses of various

* The work described here was supported in part by the National Science Foundation, the Information Systems and Information Technology Offices of DARPA under contract numbers DABT63-95-C-007 and N66001-99-D-8503, and also in part by ONR grants: N00014-95-1-1164, N00014-99-1-0377, and N00014-99-1-0380. The views presented here do not represent those of the US Government.

modalities, interfaces can be designed that can dramatically simplify human input, leading to more robust performance [5, 28]. In addition to inspiring new interface designs, such a principle can be applied at run time, whereby joint use of communication modes can compensate for errors in the individual modalities. For instance, joint use of spoken language and pen-based gesture, or speech and lip recognition, can produce better overall recognition rates than relying on the individual modes alone [23, 26, 29].

Proactive Empirical Research

Before building a complete multimodal system, it is important to understand how people would in fact interact multimodally. A series of proactive empirical studies was undertaken that investigated multimodal interaction in a variety of domains. Using high-fidelity Wizard of Oz simulations as well as actual system prototypes, it was discovered that by structuring an interface graphically, users' inputs could be channeled towards a linguistically simpler style, one that led to reduced parse ambiguity, bigram perplexity, and utterance disfluencies [2, 30], up to 2-8 fold. A number of these techniques have recently been implemented in Microsoft's MiPad prototype, and the predictions have been borne out [10].

Interface simulations investigating multimodal pen/voice interactions with map-based systems have also found that multimodal input is simpler than unimodal speech. In particular, multimodal speech and pen input to map-based applications is briefer, less syntactically complex, has fewer disfluencies, leads to fewer user errors, and is preferred over unimodal speech [25]. Furthermore, people adopt one of two styles of multimodal integration -- they gesture first, then speak (typically within 4 seconds), or they speak and gesture together, but do not speak first, gesture later [25]. Such results led directly to interface designs and to thresholds used in our QuickSet system [6]. Furthermore, the empirical simulation-based results were again borne out with QuickSet user testing.

Fusion

The core of any multimodal system is its method for fusing information derived from each mode. Depending on the characteristics of the data, information can be fused "early," at the level of signal features, and/or "late" at the level of meaning. An example of the former is audio-visual fusion, in which information about lip movement (in terms of so-called "visemes") and about the spoken words (in terms of phonemes) can be combined, resulting in better overall speech recognition, especially in noisy environments [23, 29].

In order to fuse information at the level of meaning, many groups have adopted unification of typed feature structures as the main symbolic information fusion process [8, 11, 24]. Typed feature structures are directed-acyclic attribute-value graphs, whose attributes are arranged in a type hierarchy. Such data structures are commonly used in computational linguistics to represent lexical, grammatical, and semantic information [1, 13, 22]. In the case of multimodal interaction, the meanings of the signals in each mode would be represented in such structures. Modality fusion occurs when the structures are unified [1, 11], subject to various constraints [25]. Feature structure unification, a generalization of term unification in logic programming coupled with type reasoning, is appropriate as a fusion operation because it combines complementary and redundant information, but rules out inconsistent information. In tests of QuickSet's mutual disambiguation of modalities based on feature structure unification, it was found that multimodal interaction led to a 20-40% error rate reduction over unimodal spoken language processing [26].

Fusion Architecture

Because machine perception is errorful, the software architecture needs to be designed from the ground up to handle errors. Given that recognizers produce a large number of hypotheses, with even a modest number of interpretations per mode, examining their cross-product quickly can become expensive. Thus, there needs to be rapid filtering of those combinations that could not possibly unify [14, 33], a statistical assessment of the joint probabilities of the remaining cross-modal recognition hypotheses, and finally a unification process to combine the fine structure of the interpretations. A hybrid symbolic/statistical process that can handle an arbitrary number of modes, with a variety of spatial and temporal relationships (e.g, some precede others, some co-occur, etc.) needs to be developed to arrive at the best overall interpretation of the inputs [31, 34]. Here, statistical speech and natural language processing, machine learning, pattern recognition, and sensor fusion techniques play a crucial role.

Distributed Software Architecture

The system's ability to handle parallel asynchronous input is critical to its usability. This characteristic is a significant departure from the design of current graphical user interface software, which

Figure 1: Left: QuickSet operating on a handheld PC; Right: Collaborating QuickSet operating on a 50" plasma display with touch overlay. User is speaking through a wireless microphone while drawing.

assumes that its input is certain and sequential.[1] To address these needs, we have employed a multiagent architecture [4, 15, 18] that offers fault-tolerant, distributed, asynchronous operations, with a facilitated or direct communication model.

Example 1: QuickSet

QuickSet is a collaborative handheld multimodal system based on a multiagent architecture, which controls numerous applications, including community fire and flood control, military simulators (ModSAF), exercise initialization (ExInit), and virtual terrain environments (the Naval Research Laboratory's Dragon II and SPAWAR's CommandVu). The system enables users to create point, line, and area entities on a PC screen, using a variety of form factors ranging from handheld or wearable devices to wall-sized displays, simply by speaking and sketching. For example, the user can create and position an M1A1 company at a given location and with a given orientation and posture by saying: "M1A1 company facing one two zero degrees in defensive posture," while touching the desired location. In contrast, a user of a graphical user interface (GUI) would have to locate the desired unit in a browser or palette, drag the icon onto the screen, and fill in various parameters in a dialogue box. Likewise, by speaking and sketching, the user can create linear and area features, such as unit boundaries, objectives, routes, fortifications, air corridors, no go/slow go areas, drop zones, supply routes, cultural features, etc.

QuickSet Architecture

QuickSet consists of a set of software agents, including speech recognition, natural language processing, text-to-speech, gesture recognition, multimodal integration, a map-based user interface, a database system, and an application bridge. Continuous speech and continuous gesture are processed in parallel, with n-best recognition results from each mode represented as typed feature structures. After parsing, the resulting interpretations are collected by the multimodal integrator. The integrator operates as a multimodal chart parser [12], storing partial feature structure interpretations in a multimodal chart, which is operated upon by rules, and subjected to constraints. The basic fusion operation is unification of feature structures [11].

These modality agents communicate through a central facilitator via a common language, currently the Interagent Communication Language for the Open Agent Architecture [4, 18]. The presence of the facilitator enables agents to connect,disconnect, and reconnect without restarting the system. However, as a central component, the facilitator can be a bottleneck to high bandwidth information transfer, and potentially a single point of failure. To overcome these problems, we have developed a successor multiagent system, the Adaptive Agent Architecture, which offers direct as well as facilitated communication, and supports fault-tolerant operation based on the theory of teamwork [7, 15, 16]. Importantly, the agents can reside on a variety of different types of machines, located anywhere on the Internet. If appropriately time-synchronized, distributed agents can participate in analyzing users' multimodal inputs. In particular, handheld systems, such as PDAs, can run the interface, but offload computationally intensive multimodal processing to servers operating elsewhere.

Evaluation

Multimodal interaction with QuickSet was recently compared with interaction via a standard graphical user interface (GUI) for the task of placing entities

[1] To see this, try moving the mouse on your keyboard while typing.

on a map [3]. It was found that multimodal interaction led to a 4-9 fold increase in the speed with which military users could create entities, lines and areas of various types. Although there were no more errors with multimodal interaction than with the GUI, the time to correct multimodal errors was again 4-fold faster than the time needed to correct GUI errors. Furthermore, all the users preferred interacting multimodally to using the GUI. This is just one study, comparing one GUI with multimodal interaction, but it is indicative of the potential this style of interface can offer.

Paper too?

One virtue of employing a distributed multiagent architecture as the core software architecture is that it supports a variety of platforms and hardware configurations. In particular, it is appropriate for paper-based environments.

To understand why paper is important, consider the scene below (Figure 2) from a US Army Division command post taken at a frenetic time during an exercise. On the computer screens shown here, and on the 17 other screens arrayed around the room, are military command and control systems. Here is a quiz: What's missing from this picture?

Figure 2: Scene from a US Army Division command post during an exercise (photo courtesy of William Scherlis).

Indeed, *no one* is using these or any of the other 17 systems. What the officers are in fact doing can be found in the next photo (Figure 3). They are standing on chairs, plotting the positions of military units using an 8-foot by 6-foot paper map and Post-it notes. You will notice that *they have turned their backs on computer science and technology.*

In general, it is fair to say that despite the best efforts of researchers and numerous well-funded development efforts, many military users persist in employing paper rather than computers.[2] However, the officers are not simply trying to be difficult or overly conservative in rejecting digital systems. Rather, the systems they have received are missing qualities that they value highly. The users tell us that they continue to use paper maps because they have extremely high resolution, are malleable, cheap, lightweight, and can be rolled up and taken anywhere. Importantly, *paper does not fail*, and it supports face-to-face collaboration among the staff members. For example, Figure 4 illustrates the kinds of collaboration officers engage in with paper maps, which are simply unsupported by present day computer systems.

Figure 3: Officers tracking units with paper maps and Post-it notes in preference to computer systems (Photo courtesy of William Scherlis)

[2]They are not alone in this preference. It has been observed that medical personnel in emergency and intensive care facilities, as well as air traffic controllers, prefer to interact with physical objects, such as paper and pencil, rather than use a computer [9, 17]. What is common among all these environments is their life-and-death nature, and the users' absolute requirement for safety and robustness.

Figure 4: Multiuser collaboration around a command post map.

Thus, for a variety of factors, officers (as well as medical and air traffic control personnel) prefer paper to digital systems. *We believe there is no reason they cannot have the benefits of both*.

To provide such advantages, we have adapted the QuickSet system to use paper-based interfaces, thereby enabling officers to employ their highly practiced mode of operation when using digital systems [19-21]. The new Rasa system, developed for David McGee's Ph.D. thesis, provides both sets of benefits without substantial task overhead in virtue of its understanding the symbology drawn on the paper maps and Post-it notes, as well as its understanding of the spoken language used in creating and naming units. Essentially, both paper maps and Post-it notes rest on touch or pen-sensitive digitizers. Digital ink used in drawing symbols on a Post-it note is routed to the QuickSet symbology recognizer, which creates an n-best list of hypothesized units that the user drew. Accompanying speech is recognized and parsed, and the fused interpretation then waits at the integrator for a location feature to arrive. The act of placing the note on the map, which is overlaid on a touch-sensitive digitizer, then supplies the desired location. It is most important to note that in this tangible multimodal system, the physical artifacts, i.e., the paper map and Post-Its, *become* the computational interface. For example, moving a Post-it note on the map moves it in the digital system. In response, Rasa projects system updates onto the paper map. Rasa can also provide data to other displays and visualizations.

The system has been evaluated with military users, and has been found to be as fast as paper, and is preferred to paper [21]. Moreover, users' work continues when the system or computer communications fail, and the effort to synchronize them when the system is brought on line is well within users' tolerance. Clearly, given the situation shown in Figure 3, we would hypothesize that it would offer superior usability to the existing command-and-control systems**.** This hypothesis will be tested in field experiments.

Concluding Remarks

With this very brief paper, I hope to have called attention through the QuickSet and Rasa examples to the potential for a variety of subfields within AI and computer science (e.g., speech, natural language, vision, distributed systems, machine learning, knowledge-representation and reasoning, human-computer interaction), in collaboration with numerous other disciplines (cognitive science, ethnography, linguistics, pattern recognition, sensor fusion, etc.), to contribute to radically changing the human-computer interface. In the development and evaluation of QuickSet, we see a direct progression from proactive empirical research, to system development, and finally to formal laboratory and field user testing. For Rasa, our taking an ethnographic perspective and observing the actual "work practice," enabled us to identify both problems and opportunities for technology. In a very real sense, neither of these systems could have been developed without multidisciplinary collaboration. No one methodology was employed, nor could it have been. The research within each "core" discipline was focused by the multidisciplinary goals to produce a synergistic whole.

Other domains ripe for multimodal interaction include in-home access to digital information, mobile computing, geographic information systems (GIS), computer-aided design (CAD), games. Overall, multimodal interaction can benefit society by enlarging the base of users to children, users with disabilities, or users whose physical situations of computer usage are changing. A concentrated effort to research and develop such systems can have enormous scientific and technological payoffs, and would be a worthy complement to the other focal areas of AI research.

Acknowledgements

Many thanks to my colleagues at CHCC/OGI, who made this research possible, including (but not limited to): Josh Clow, Marcus Huber, Michael Johnston, Sanjeev Kumar, David McGee, James Pittman, Sharon Oviatt, Ira Smith, Matt Wesson, and Lizhong Wu.

References

1. Carpenter, R., *The logic of typed feature structures*. 1992, Cambridge University Press: Cambridge,England.
2. Cohen, P.R., *The role of natural language in a multimodal interface*, in *Proceedings of the Conference on User Interface Software Technology (UIST'92)*. 1992, ACM Press: New York.
3. Cohen, P.R., McGee, D., and Clow, J., The efficiency of multimodal interaction for a map-based task, in the *Proceedings of the Proceedings of the 6th Applied Natural Language Processing Conference*, Seattle, Washington, May, 2000, Association for Computational Linguistics, 331-338.
4. Cohen, P.R., Cheyer, A., Wang, M.Q., *et al.*, *An Open Agent Architecture*, in *Working notes of the AAAI Spring Symposium Series on Software Agents*. 1994, AAAI Press, reprinted in Readings in Agents, Huhns and Singh (eds.), Morgan Kaufmann Publishers, Inc., San Mateo, California, 1997: Stanford, Calif. 1-8.
5. Cohen, P.R., Dalrymple, M., Moran, D.B., *et al.*, *Synergistic use of natural language and direct manipulation*, in *Proceedings of the Human-Factors in Computing Systems Conference (CHI'89)*. 1989, ACM Press: New York.
6. Cohen, P.R., Johnston, M., McGee, D., *et al.*, *QuickSet: Multimodal interaction for distributed applications*, in *Proceeedings of the Fifth ACM International Multmedia Conference*, E. Glinert, Editor. 1997, ACM Press: New York. 31-40.
7. Cohen, P.R. and Levesque, H.J., *Teamwork.* Nous, 1991. **25**(4): 487-512.
8. Denecke, M., and Yang, J., Partial information in multimodal dialogue, in the *Proceedings of the Proceedings of the International Conference on Multimodal Interaction*, Beijing, China, 2000, 624-633.
9. Gorman, P., Ash, J., Lavelle, M., Lyman, J., Delcambre, L., and Maier, D., *Bundles in the Wild: Managing information to solve problems and maintain situation awareness.* Library Trends, 2000. **49**(2).
10. Huang, X., Acero, A., Chelba, C., Deng, L., Duchene, D., Goodman, J., Hon, H., Jacoby, D., Jiang, L., Loynd, R., Mahajan, M., Mau, P., Meredith, S., Mughal, S., Neto, S., Plumpe, M., Wang, K., Wang, Y., MIPAD: A next generation PDA prototype, in the *Proceedings of the Proceedings of the International Conference on Spoken Language Processing*, Beijing, China, 2000, Chinese Friendship Publishers.
11. Johnston, M., P. R. Cohen, D. McGee, S. L. Oviatt, J. A. Pittman, I. Smith., Unification-based multimodal integration., in the *Proceedings of the Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and 8th Conference of the European Chapter of the Association for Computational Linguistics*, 1997, 281-288.
12. Johnston, M., Unification-based Multimodal Parsing, in the *Proceedings of the Proceedings of COLING-ACL 98: The 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics*, 1998, Association for Computational Linguistics.
13. Kay, M., *Functional Grammar*, in *Proceedings of the Fifth Annual Meeting of the Berkeley Linguistics Society*. 1979: Berkeley, California. 142-158.
14. Kiefer, B.K., H.-U., Nederhof, M.-J., *Efficient and robust parsing of word hypotheses graphs*, in *Verbmobil: Foundations of speech-to-speech translation*, W. Wahlster, Editor. 2000, Springer: Berlin, Germany. 280-295.
15. Kumar, S., Cohen, P. R., Towards a Fault-Tolerant Multi-Agent System Architecture., in the *Proceedings of the Proceedings of The Fourth International Conference on Autonomous Agents (Agents 2000)*, Barcelona, Spain, 2000, AAAI Press, 459-466.
16. Levesque, H.J., Cohen, P.R., and Nunes, J., *On Acting Together*, in *Proceedings of American Association for Artificial Intelligence (AAAI-90)*. 1990: San Mateo, California.
17. Mackay, W.E., *Is paper safer? The role of flight strips in air traffic control.* ACM Transactions on Human-Computer Interaction, 1999. **6**(4): 311-340.
18. Martin, D.L., Cheyer, A., J., Moran, D. B., *The Open Agent Architecture: A framework for building distributed software systems.* Applied Artificial Intelligence, 1999. **13**(January-March): 91-128.
19. McGee, D., Cohen, P. R., and Wu, L., Something from nothing: Augmenting a paper-based work practice via multimodal interaction, in the *Proceedings of the Proceedings of the Conference on Design of Augmented Reality Environments*, Denmark, 2000, Association for Computing Machinery.
20. McGee, D., Cohen, P. R., Creating Tangible Interfaces by Augmenting Physical Objects with

Multimodal Language, in the *Proceedings of the Proceedings of International Conference on Intelligent User Interfaces*, Santa Fe, New Mexico, 2001.

21. McGee, D.R., *Augmenting environments with multimodal interaction*, in *Dept. of Computer Science and Engineering*. forthcoming, Oregon Graduate Institute of Science and Technology: Beaverton, OR.
22. McKeown, K.R., *Generating natural language text in response to questions about database structure.* 1982.
23. Neti, C., Iyengar, G., Potamianos, G., Senior, A., Perceptual interface for information interaction: Joint processing of audio and visual information for human-computer interaction, in the *Proceedings of the Proceedings of the International Conference on Spoken Language Processing (ICSLP'2000)*, Beijing, China, 2000, Chinese Friendship Publishers, 11-14.
24. Nigay, L., and Coutaz, J., A generic platform for addressing the multimodal challenge, in the *Proceedings of the Proceedings of the Conference on Human Factors in Computing Systems (CHI'95)*, 1995, ACM Press, 98-105.
25. Oviatt, S.L., *Multimodal interactive maps: Designing for human performance.* Human Computer Interaction, 1997. **12**: 93-129.
26. Oviatt, S.L., Mutual disambiguation of recognition errors in a multimodal architecture, in the *Proceedings of the Human Factors in Computing Systems (CHI'99)*, New York, 1999, ACM Press, 576-583.
27. Oviatt, S.L., Cohen, P. R., *Multimodal interfaces that process what comes naturally.* Communications of the ACM, 2000. **Volume 43**(3): 45-49.
28. Oviatt, S.L., *Taming recognition errors with a multimodal interface.* Communications of the ACM, 2000. **43**(9): 45-51.
29. Oviatt, S.L., Cohen, P. R., Wu, L., Vergo, J., Duncan, L., Suhm, B., Bers, J., Holzman, T., Winograd, T., Landay, J., Larson, J., Ferro, D., *Designing the user interface for multimodal speech and gesture applications: State-of-the-art systems and research directions for 2000 and beyond.* Human Computer Interaction, 2001. **15**(4): 263-322.
30. Oviatt, S.L., Cohen, P.R., and Wang, M.Q., *Toward Interface Design for Human Language Technology: Modality and Structure as Determinants of Linguistic Complexity.* Speech Communication, 1994. **15**(3-4).
32. Stork, D.G., Wolff, G., and Levine, E., *Neural network lipreading system for improved speech recognition*, in *Proceedings of the International Joint Conference on Neural Networks Vol. II*. 1992. 286-295.
33. Suhm, B., Myers, B., and Waibel, A., Model-based and empirical evaluation of multimodal interactive error correction, in the *Proceedings of the Human Factors in Computing Systems (CHI'99)*, New York, 1999, ACM Press, 584-591.
34. Wu, L., , Oviatt, S., L. and Cohen, P. R.,, *Statistical multimodal integration for intelligent HCI*, in *Neural Networks for Signal Processing*, Y.H. Hu, Larsen, J., Wilson, E., and Douglas, S., Editor. 1999, IEEE Press: New York. 487-496.

Plausibility Measures: A General Approach For Representing Uncertainty

Joseph Y. Halpern*
Cornell University
Dept. of Computer Science
Ithaca, NY 14853
halpern@cs.cornell.edu
http://www.cs.cornell.edu/home/halpern

1 Introduction

The standard approach to modeling uncertainty is probability theory. In recent years, researchers, motivated by varying concerns including a dissatisfaction with some of the axioms of probability and a desire to represent information more qualitatively, have introduced a number of generalizations and alternatives to probability, including Dempster-Shafer belief functions [Shafer, 1976], possibility measures [Dubois and Prade, 1990], lexicographic probability [Blume *et al.*, 1991], and many others. Rather than investigating each of these approaches piecemeal, I consider here an approach to representing uncertainty that generalizes them all, and lets us understand their commonalities and differences.

A *plausibility measure* [Friedman and Halpern, 1995] associates with a set a *plausibility*, which is just an element in a partially ordered space. The only real requirement is that if U is a subset of V, then the plausibility of U is less than equal to the plausibility of V. Probability measures are clearly plausibility measures; every other representation of uncertainty that I am aware of can also be viewed as a plausibility measure. Given how little structure plausibility measures have, it is perhaps not surprising that plausibility measures generalize so many other notions. This very lack of structure turns out to be a significant advantage. By adding structure on an "as needed" basis, it is possible to characterize what is required to ensure that a plausibility measure has certain properties of interest. This both gives insight into the essential features of the properties in question and makes it possible to prove general results that apply to many representations of uncertainty.

In this paper, I discuss three examples of this phenomenon.

- belief, belief revision, and default reasoning,
- expectation and decision making,
- compact representations of uncertainty (Bayesian networks).

Most of the discussion is based on earlier work (some of it joint with Nir Friedman). In the next two sections I define plausibility measures and conditional plausibility measures. The next three sections considers each of the topics above in more detail.

*Supported in part by NSF under grants IRI-96-25901 and IIS-0090145.

2 Plausibility Measures

A probability space is a tuple $(W, \mathcal{F}, \mu)$, where W is a set of worlds, $\mathcal{F}$ is an algebra of *measurable* subsets of W (that is, a set of subsets closed under union and complementation to which we assign probability), and μ is a *probability measure*, that is, a function mapping each set in $\mathcal{F}$ to a number in $[0, 1]$ satisfying the well-known Kolmogorov axioms ($\mu(\emptyset) = 0$, $\mu(W) = 1$, and $\mu(U \cup V) = \mu(U) + \mu(V)$ if U and V are disjoint).[1]

A plausibility space is a direct generalization of a probability space. Simply replace the probability measure μ by a *plausibility measure* Pl that, rather than mapping sets in $\mathcal{F}$ to numbers in $[0, 1]$, maps them to elements in some arbitrary partially ordered set. $\mathrm{Pl}(U)$ is read "the plausibility of set U". If $\mathrm{Pl}(U) \leq \mathrm{Pl}(V)$, then V is at least as plausible as U. Formally, a *plausibility space* is a tuple $S = (W, \mathcal{F}, \mathrm{Pl})$, where W is a set of worlds, $\mathcal{F}$ is an algebra over W, and Pl maps sets in $\mathcal{F}$ to some set D of *plausibility values* partially ordered by a relation $\leq_D$ (so that $\leq_D$ is reflexive, transitive, and anti-symmetric). D is assumed to include two special elements, $\top_D$ and $\bot_D$, such that $\bot_D \leq_D d \leq_D \top_D$ for all $d \in D$. In the case of probability measures, $D = [0, 1]$, and $\top_D$ and $\bot_D$ are 1 and 0, respectively. As usual, the ordering $<_D$ is defined by taking $d_1 <_D d_2$ if $d_1 \leq_D d_2$ and $d_1 \neq d_2$. I omit the subscript D from $\leq_D$, $<_D$, $\top_D$, and $\bot_D$ whenever it is clear from context.

There are three requirements on plausibility measures. The first two are analogues of the conventions that hold for all representations of uncertainty: the whole space gets the maximum plausibility and the empty set gets the minimum plausibility. The third requirement says that a set must be at least as plausible as any of its subsets.

Pl1. $\mathrm{Pl}(W) = \top$.

Pl2. $\mathrm{Pl}(\emptyset) = \bot$.

Pl3. If $U \subseteq V$, then $\mathrm{Pl}(U) \leq \mathrm{Pl}(V)$.

Since $\leq_D$ is a partial order, Pl3 says that, if $U \subseteq V$, then the plausibility of U is comparable to the plausibility of V and, moreover, $\mathrm{Pl}(U) \leq \mathrm{Pl}(V)$.

[1] Frequently it is also assumed that μ is *countably additive*, i.e., if U_i, $i > 0$, are pairwise disjoint, then $\mu(\bigcup_i U_i) = \sum_i \mu(U_i)$. Since I focus on finite state spaces here, countable additivity does not play a significant role, so I do not assume it.

Clearly probability spaces are instances of plausibility spaces. Almost all other representations of uncertainty in the literature can also be viewed as instances of plausibility measures. Here are some examples:

- A *belief function* on W is a function $\mathrm{Bel} : 2^W \to [0,1]$ satisfying certain axioms [Shafer, 1976]. These axioms certainly imply property Pl3, so a belief function is a plausibility measure. There is a corresponding *plausibility function* Plaus defined as $\mathrm{Plaus}(U) = 1 - \mathrm{Bel}(\overline{U})$.[2]
- A *possibility measure* [Dubois and Prade, 1990] on W is a function $\mathrm{Poss} : 2^W \to [0,1]$ such that $\mathrm{Poss}(W) = 1$, $\mathrm{Poss}(\emptyset) = 0$, and $\mathrm{Poss}(U) = \sup_{w \in U}(\mathrm{Poss}(\{w\}))$.
- An *ordinal ranking* (or κ-*ranking*) κ on W (as defined by [Goldszmidt and Pearl, 1992], based on ideas that go back to [Spohn, 1988]) is a function mapping subsets of W to $\mathbb{N}^* = \mathbb{N} \cup \{\infty\}$ such that $\kappa(W) = 0$, $\kappa(\emptyset) = \infty$, and $\kappa(U) = \min_{w \in U}(\kappa(\{w\}))$. Intuitively, an ordinal ranking assigns a degree of surprise to each subset of worlds in W, where 0 means unsurprising and higher numbers denote greater surprise. It is easy to see that a ranking κ is a plausibility measure with range $\mathbb{N}$, where $x \leq_{\mathbb{N}^*} y$ if and only if $y \leq x$ under the usual ordering.
- A *lexicographic probability system* (LPS) [Blume *et al.*, 1991] of length m is a sequence $\vec{\mu} = (\mu_0, \ldots, \mu_m)$ of probability measures. Intuitively, the first measure in the sequence, μ_0, is the most important one, followed by μ_1, μ_2, and so on. Very roughly speaking, the probability assigned to an event U by a sequence such as (μ_0, μ_1) can be taken to be $\mu_0(U) + \epsilon\mu_1(U)$, where ϵ is an infinitesimal. Thus, even if the probability of U according to μ_0 is 0, U still has a positive (although infinitesimal) probability if $\mu_1(U) > 0$.

In all these cases, the plausibility ordering is total. But there are also cases of interest where the plausibility ordering is *not* total. For example, suppose that $\mathcal{P}$ is a set of probability measures on W. Let $\mathcal{P}_*$ be the lower probability of $\mathcal{P}$, so that $\mathcal{P}_*(U) = \inf\{\mu(U) : \mu \in \mathcal{P}\}$. Similarly, the upper probability $\mathcal{P}^*$ is defined as $\mathcal{P}^*(U) = \sup\{\mu(U) : \mu \in \mathcal{P}\}$.

Both $\mathcal{P}_*$ and $\mathcal{P}^*$ give a way of comparing the likelihood of two subsets U and V of W. These two ways are incomparable: it is easy to find a set $\mathcal{P}$ of probability measures on W and subsets U and V of W such that $\mathcal{P}_*(U) < \mathcal{P}_*(V)$ and $\mathcal{P}^*(U) > \mathcal{P}^*(V)$. Rather than choosing between $\mathcal{P}_*$ and $\mathcal{P}^*$, we can associate a different plausibility measure with $\mathcal{P}$ that captures both. Let $D_{\mathcal{P}_*,\mathcal{P}^*} = \{(a,b) : 0 \leq a \leq b \leq 1\}$ and define $(a,b) \leq (a',b')$ iff $b \leq a'$. This puts a partial order on $D_{\mathcal{P}_*,\mathcal{P}^*}$, with $\bot_{D_{\mathcal{P}_*,\mathcal{P}^*}} = (0,0)$ and $\top_{D_{\mathcal{P}_*,\mathcal{P}^*}} = (1,1)$. Define $\mathrm{Pl}_{\mathcal{P}_*,\mathcal{P}^*}(U) = (\mathcal{P}_*(U), \mathcal{P}^*(U))$. Thus, $\mathrm{Pl}_{\mathcal{P}_*,\mathcal{P}^*}$ associates with a set U two numbers that can be thought of as defining an interval in terms of the lower and upper probability of U. It is easy to check that $\mathrm{Pl}_{\mathcal{P}_*,\mathcal{P}^*}(U) \leq \mathrm{Pl}_{\mathcal{P}_*,\mathcal{P}^*}(V)$ if the upper probability of U is less than or equal to the lower probability of V. Clearly, $\mathrm{Pl}_{\mathcal{P}_*,\mathcal{P}^*}$ satisfies Pl1–3, so it is indeed a plausibility measure, but one that puts only a partial (pre)order on events. A similar plausibility measure can be associated with a belief/plausibility function.

The trouble with $\mathcal{P}_*$, $\mathcal{P}^*$, and even $\mathrm{Pl}_{\mathcal{P}_*,\mathcal{P}^*}$ is that they lose information. For example, it is not hard to find a set $\mathcal{P}$ of probability measures and subsets U, V of W such that $\mu(U) \leq \mu(V)$ for all $\mu \in \mathcal{P}$ and $\mu(U) < \mu(V)$ for some $\mu \in \mathcal{P}$, but $\mathcal{P}_*(U) = \mathcal{P}_*(V)$ and $\mathcal{P}^*(U) = \mathcal{P}^*(V)$. Indeed, there exists an infinite set $\mathcal{P}$ of probability measures such that $\mu(U) < \mu(V)$ for all $\mu \in \mathcal{P}$ but $\mathcal{P}_*(U) = \mathcal{P}_*(V)$ and $\mathcal{P}^*(U) = \mathcal{P}^*(V)$. If all the probability measures in $\mathcal{P}$ agree that U is less likely than V, it seems reasonable to conclude that U is less likely than V. However, none of $\mathcal{P}_*$, $\mathcal{P}^*$, or $\mathrm{Pl}_{\mathcal{P}_*,\mathcal{P}^*}$ necessarily draw this conclusion.

It is not hard to associate yet another plausibility measure with $\mathcal{P}$ that does not lose this important information (and does indeed conclude that U is less likely than V). Suppose, without loss of generality, that there is some index set I such that $\mathcal{P} = \{\mu_i : i \in I\}$. Thus, for example, if $\mathcal{P} = \{\mu_1, \ldots, \mu_n\}$, then $I = \{1, \ldots, n\}$. (In general, I may be infinite.) Let D_I consist of all functions from I to $[0,1]$. The standard pointwise ordering on functions—that is, $f \leq g$ if $f(i) \leq g(i)$ for all $i \in I$—gives a partial order on D_I. Note that $\bot_{D_I}$ is the function $f : I \to [0,1]$ such that $f(i) = 0$ for all $i \in I$ and $\top_{D_I}$ is the function g such that $g(i) = 1$ for all $i \in I$. For $U \subseteq W$, let f_U be the function such that $f_U(i) = \mu_i(U)$ for all $i \in I$. Define the plausibility measure $\mathrm{Pl}_{\mathcal{P}}$ by taking $\mathrm{Pl}_{\mathcal{P}}(U) = f_U$. Thus, $\mathrm{Pl}_{\mathcal{P}}(U) \leq \mathrm{Pl}_{\mathcal{P}}(V)$ iff $f_U(i) \leq f_V(i)$ for all $i \in I$ iff $\mu(U) \leq \mu(V)$ for all $\mu \in \mathcal{P}$. It is easy to see that $f_\emptyset = \bot_{D_I}$ and $f_W = \top_{D_I}$. Clearly $\mathrm{Pl}_{\mathcal{P}}$ satisfies Pl1–3. Pl1 and Pl2 follow since $\mathrm{Pl}_{\mathcal{P}}(\emptyset) = f_\emptyset = \bot_{D_I}$ and $\mathrm{Pl}_{\mathcal{P}}(W) = f_W = \top_{D_I}$, while Pl3 holds because if $U \subseteq V$, then $\mu(U) \leq \mu(V)$ for all $\mu \in \mathcal{P}$.

To see how this representation works, consider a simple example where a coin which is known to be either fair or double-headed is tossed. The uncertainty can be represented by two probability measures on μ_1, which gives heads probability 1, and μ_2 which gives heads probability $1/2$. Taking the index set to be $\{1,2\}$, this gives us a plausibility measure $\mathrm{Pl}_{\mathcal{P}}$ such that $\mathrm{Pl}_{\mathcal{P}}(H)$ is a function f such that $f(1) = 1$ and $f(2) = 1/2$; similarly, $\mathrm{Pl}_{\mathcal{P}}(T)$ is a function f' such that $f'(1) = 0$ and $f'(2) = 1/2$.

3 Conditional Plausibility

Suppose an agent's beliefs are represented by a plausibility measure Pl. How should these beliefs be updated in light of new information? The standard approach to updating in probability theory is by conditioning. Most other representations of uncertainty have an analogue to conditioning. Indeed, compelling arguments have been made in the context of probability to take conditional probability as a primitive notion, rather than unconditional probability. The idea is to start with a primitive notion $\Pr(\cdot|\cdot)$ satisfying some constraints (such as $\Pr(U \cup U'|V) = \Pr(U|V) + \Pr(U'|V)$ if U and U' are disjoint) rather than starting with an unconditional probability measure and defining conditioning in terms of it. The advantage of taking conditional probability as primitive is that it

[2] The word "plausibility" is slightly overloaded, appearing both in the context of "plausibility function" and "plausibility measure". Plausibility functions will play only a minor role in this paper, so there should not be much risk of confusion.

allows conditioning on events of unconditional probability 0. (If W is the whole space, the unconditional probability of V can be identified with $\Pr(V|W)$; note that $\Pr(U|V)$ may be well defined even if $\Pr(V|W) = 0$.) Although conditioning on events of measure 0 may seem to be of little practical interest, it turns out to play a critical role in game theory (see, for example, [Blume *et al.*, 1991; Myerson, 1986]), the analysis of conditional statements (see [Adams, 1966; McGee, 1994]), and in dealing with nonmonotonicity (see, for example, [Lehmann and Magidor, 1992]).

Most other representations of uncertainty also have an associated notion of conditioning. I now discuss a notion of conditional plausibility that generalizes them all. A *conditional plausibility measure (cpm)* maps pairs of subsets of W to some partially ordered set D. I write $\mathrm{Pl}(U|V)$ rather than $\mathrm{Pl}(U, V)$, in keeping with standard notation. An important issue in defining conditional plausibility is to make precise what the allowable arguments to Pl are. I take the domain of a cpm to have the form $\mathcal{F} \times \mathcal{F}'$ where, roughly speaking, $\mathcal{F}'$ consists of those sets in $\mathcal{F}$ on which conditioning is allowed. For example, for a conditional probability measure defined in the usual way from an unconditional probability measure μ, $\mathcal{F}'$ consists of all sets V such that $\mu(V) > 0$. (Note that $\mathcal{F}'$ is not an algebra—it is not closed under complementation.) A *Popper algebra* over W is a set $\mathcal{F} \times \mathcal{F}'$ of subsets of $W \times W$ satisfying the following properties:

Acc1. $\mathcal{F}$ is an algebra over W.

Acc2. $\mathcal{F}'$ is a nonempty subset of $\mathcal{F}$.

Acc3. $\mathcal{F}'$ is closed under supersets in $\mathcal{F}$; that is, if $V \in \mathcal{F}'$, $V \subseteq V'$, and $V' \in \mathcal{F}$, then $V' \in \mathcal{F}'$.

(Popper algebras are named after Karl Popper, who was the first to consider formally conditional probability as the basic notion [Popper, 1968]. This definition of cpm is from [Halpern, 2000a] which in turn is based on the definition in [Friedman and Halpern, 1995].)

A *conditional plausibility space* (*cps*) is a tuple $(W, \mathcal{F}, \mathcal{F}', \mathrm{Pl})$, where $\mathcal{F} \times \mathcal{F}'$ is a Popper algebra over W, $\mathrm{Pl} : \mathcal{F} \times \mathcal{F}' \to D$, D is a partially ordered set of plausibility values, and Pl is a *conditional plausibility measure* (cpm) that satisfies the following conditions:

CPl1. $\mathrm{Pl}(\emptyset|V) = \bot$.

CPl2. $\mathrm{Pl}(W|V) = \top$.

CPl3. If $U \subseteq U'$, then $\mathrm{Pl}(U|V) \leq \mathrm{Pl}(U'|V)$.

CPl4 $\mathrm{Pl}(U|V) = \mathrm{Pl}(U \cap V|V)$.

CPl1–3 are the obvious analogues to Pl1–3. CPl4 is a minimal property that guarantees that when conditioning on V, everything is relativized to V. It follows easily from CPl1–4 that $\mathrm{Pl}(\cdot|V)$ is a plausibility measure on V for each fixed V. A cps is *acceptable* if it satisfies

Acc4. If $V \in \mathcal{F}'$, $U \in \mathcal{F}$, and $\mathrm{Pl}(U|V) \neq \bot$, then $U \cap V \in \mathcal{F}'$.

Acceptability is a generalization of the observation that if $\Pr(V) \neq 0$, then conditioning on V should be defined. It says that if $\mathrm{Pl}(U|V) \neq \bot_D$, then conditioning on $V \cap U$ should be defined. A cps $(W, \mathcal{F}, \mathcal{F}', \mathrm{Pl})$ is *standard* if $\mathcal{F}' = \{U : \mathrm{Pl}(U|W) \neq \bot\}$.

CPl1–4 are rather minimal requirements. For example, they do not place any constraints on the relationship between $\mathrm{Pl}(U|V)$ and $\mathrm{Pl}(U|V')$ if $V \neq V'$. One natural additional condition is the following.

CPl5. If $V \cap V' \in \mathcal{F}'$ and $U, U' \in \mathcal{F}$, then $\mathrm{Pl}(U|V \cap V') \leq \mathrm{Pl}(U'|V \cap V')$ iff $\mathrm{Pl}(U \cap V|V') \leq \mathrm{Pl}(U' \cap V|V')$.

It is not hard to show that CPl5 implies CPl4. While it seems reasonable, note that CPl5 does not hold in some cases of interest. For example, there are two well-known ways of defining conditioning for belief functions (see [Halpern and Fagin, 1992]), one using Dempster's rule of combination and the other treating belief functions as lower probabilities. They both satisfy CPl1–4, and neither satisfies CPl5.

Many plausibility spaces of interest have more structure. In particular, there are analogues to addition and multiplication. More precisely, there is a way of computing the plausibility of the union of two disjoint sets in terms of the plausibility of the individual sets and a way of computing $\mathrm{Pl}(U \cap V|V')$ given $\mathrm{Pl}(U|V \cap V')$ and $\mathrm{Pl}(V|V')$. A cps $(W, \mathcal{F}, \mathcal{F}', \mathrm{Pl})$ where Pl has range D is *algebraic* if it is acceptable and there are functions $\oplus : D \times D \to D$ and $\otimes : D \times D \to D$ such that the following properties hold:

Alg1. If $U, U' \in \mathcal{F}$ are disjoint and $V \in \mathcal{F}'$ then $\mathrm{Pl}(U \cup U'|V) = \mathrm{Pl}(U|V) \oplus \mathrm{Pl}(U'|V)$.

Alg2. If $U \in \mathcal{F}$, $V \cap V' \in \mathcal{F}'$, then $\mathrm{Pl}(U \cap V|V') = \mathrm{Pl}(U|V \cap V') \otimes \mathrm{Pl}(V|V')$.

Alg3. $\otimes$ distributes over $\oplus$; more precisely, $a \otimes (b_1 \oplus \cdots \oplus b_n) = (a \otimes b_1) \oplus \cdots \oplus (a \otimes b_n)$ if $(a, b_1), \ldots, (a, b_n), (a, b_1 \oplus \cdots \oplus b_n) \in \mathit{Dom}_{\mathrm{Pl}}(\otimes)$ and $(b_1, \ldots, b_n), (a \otimes b_1, \ldots, a \otimes b_n) \in \mathit{Dom}_{\mathrm{Pl}}(\oplus)$, where $\mathit{Dom}_{\mathrm{Pl}}(\oplus) = \{(\mathrm{Pl}(U_1|V), \ldots, \mathrm{Pl}(U_n|V)) : U_1, \ldots, U_n \in \mathcal{F}$ are pairwise disjoint and $V \in \mathcal{F}'\}$ and $\mathit{Dom}_{\mathrm{Pl}}(\otimes) = \{(\mathrm{Pl}(U|V \cap V'), \mathrm{Pl}(V|V')) : U \in \mathcal{F}, V \cap V' \in F'\}$.

Alg4. If $(a, c), (b, c) \in \mathit{Dom}_{\mathrm{Pl}}(\otimes)$, $a \otimes c \leq b \otimes c$, and $c \neq \bot$, then $a \leq b$.

I sometimes refer to the cpm Pl as being algebraic as well.

There are well-known techniques for extending some standard unconditional representations of uncertainty to conditional representations. All satisfy CPl1-4, when viewed as plausibility measures. (Indeed, as shown in [Halpern, 2000a], there is a construction for converting an arbitrary unconditional plausibility space $(W, \mathcal{F}, \mathrm{Pl})$ to an acceptable standard cps.) In many cases, the resulting cps is algebraic. But one important case that is not algebraic is conditional belief functions (using either definition of conditioning).

To give one example of a construction that does lead to an algebraic cps, consider LPS's. Blume, Brandenburger, and Dekel 1991 (BBD) define conditioning in LPS's as follows. Given $\vec{\mu}$ and $U \in \mathcal{F}$ such that $\mu_i(U) > 0$ for some index i, let $\vec{\mu}|V = (\mu_{k_0}(\cdot|V), \ldots, \mu_{k_m}(\cdot|V))$, where $(k_0, \ldots, k_m)$ is the subsequence of all indices for which the probability of U is positive. Thus, the length of the LPS $\vec{\mu}|V$ depends on V. Let D^k consist of all sequences $(a_0, \ldots, a_k) \notin \{(0, \ldots, 0), (1, \ldots, 1)\}$ such that $a_i \in [0, 1]$ for $i = 0, \ldots, k$, and let $D = \{\mathbf{0}, \mathbf{1}\} \cup (\cup_{k=0}^{\infty} D_k)$. Roughly speaking, $\mathbf{0}$ is

meant to represent all sequences of the form $(0, \ldots, 0)$, whatever their length; similarly, $\mathbf{1}$ represents all sequences of the form $(1, \ldots, 1)$. Define a partial order $\leq_D$ on D so that $d_1 \leq_D d_2$ if $d_1 = \mathbf{0}$, $d_2 = \mathbf{1}$, or d_1 and d_2 are vectors of the same length and d_1 is lexicographically less than or equal to d_2. Note that vectors of different length are incomparable.

An unconditional LPS $\vec{\mu}$ defined on an algebra $\mathcal{F}$ over W can then be extended to a standard cps $(W, \mathcal{F}, \mathcal{F}', \vec{\mu})$ using the definition of conditioning above. Note that although $\vec{\mu}(U|V)$ may be incomparable to $\vec{\mu}(U'|V')$ for $V \neq V'$, $\vec{\mu}(U|V)$ will definitely be comparable to $\vec{\mu}(U'|V)$. Moreover, the definition of $\mathbf{0}$ and $\mathbf{1}$ guarantees that $\mathbf{0} = \vec{\mu}(\emptyset|U') \leq_D \vec{\mu}(V|U) \leq_D \vec{\mu}(U''|U'') = \mathbf{1}$ if $U', U'' \in \mathcal{F}'$, as required by CPl1 and CPl2.

The cps $(W, \mathcal{F}, \mathcal{F}', \vec{\mu})$ is in fact algebraic; $\oplus$ and $\otimes$ are functions that satisfy the following constraints:

- if d_1 and d_2 are vectors of the same length, $d_1 \oplus d_2 = d_1 + d_2$ (where $+$ represents pointwise addition),
- $d \oplus \mathbf{0} = \mathbf{0} \oplus d = d$,
- $d \otimes \mathbf{1} = \mathbf{1} \otimes d = d$,
- $\mathbf{0} \otimes d = d \otimes \mathbf{0} = \mathbf{0}$,
- $(a_1, \ldots, a_m) \otimes (\vec{0}, b_1, \vec{0}, \ldots, \vec{0}, b_m, \vec{0}) = (\vec{0}, a_1 b_1, \vec{0}, \ldots, \vec{0}, a_m b_m, \vec{0})$, where $\vec{0}$ represents a possibly empty sequence of 0s, and $b_1, \ldots, b_m > 0$.

I leave it to the reader to check that these definitions indeed make the cps algebraic.

A construction similar in spirit can be used to define a notion of conditioning appropriate for the representation $\mathrm{Pl}_{\mathcal{P}}$ of a set $\mathcal{P}$ of plausibility measures; this also leads to an algebraic cps [Halpern, 2000a].

4 Belief Revision and Default Reasoning

4.1 Belief

There have been many models used to capture belief. Perhaps the best known approach uses Kripke structures [Hintikka, 1962], where an agent believes φ if φ is true at all worlds the agent considers possible. In terms of events (sets of worlds), an agent believes U if U contains all the worlds that the agent considers possible. Another popular approach is to use probability: an agent believes U if the probability of U is at least $1 - \epsilon$ for some appropriate $\epsilon \geq 0$.

One of the standard assumptions about belief is that it is closed under conjunction: if an agent believes U_1 and U_2, then the agent should also believe $U_1 \cap U_2$. This holds for the definition in terms of Kripke structures. It holds for the probabilistic definition only if $\epsilon = 0$. Indeed, identifying knowledge/belief with "holds with probability 1" is common, especially in the economics/game theory literature [Brandenburger and Dekel, 1987].

A number of other approaches to modeling belief have been proposed recently, in the game theory and philosophy literature. One, due to Brandenburger 1999, uses *filters*. Given a set W of possible worlds, a *filter* $\mathcal{F}$ is a nonempty set of subsets of W that (1) is closed under supersets (so that if $U \in \mathcal{F}$ and $U \subseteq U'$, then $U' \in \mathcal{F}$), (2) is closed under finite intersection (so that if $U, U' \in \mathcal{F}$, then $U \cap U' \in \mathcal{F}$), and (3) does not contain the empty set. Given a filter $\mathcal{F}$, an agent is said to believe U iff $U \in \mathcal{F}$. Note that the set of sets which are given probability 1 by a probability measure form a filter. Conversely, every filter $\mathcal{F}$ defines a finitely additive probability measure Pr: the sets in $\mathcal{F}$ get probability 1; all others get probability 0. We can also obtain a filter from the Kripke structure definition of knowledge. If the agent considers possible the set $U \subseteq W$, then let $\mathcal{F}$ consist of all superset of U. This is clearly a filter (consisting of precisely the events the agent believes). Conversely, in a finite space, a filter $\mathcal{F}$ determines a Kripke structure. The agent considers possible precisely the intersection of all the sets in $\mathcal{F}$ (which is easily seen to be nonempty). In an infinite space, a filter may not determine a Kripke structure precisely because the intersection of all sets in the filter may be empty. The events believed in a Kripke structure form a filter whose sets are closed under arbitrary intersection.

Another approach to modeling belief, due to Brandenburger and Keisler 2000, uses LPS's. Say that an agent *believes U in LPS $\vec{\mu}$* if there is some $j \leq m$ such that $\mu_i(U) = 1$ for all $i \leq j$ and $\mu_i(U) = 0$ for $i > j$. It is easy to see that beliefs defined this way are closed under intersection. Brandenburger and Keisler give an elegant decision-theoretic justification for this notion of belief. Interestingly, van Fraassen 1995 defines a notion of belief using conditional probability spaces that can be shown to be closely related to the definition given by Brandenburger and Keisler.

Plausibility measures provide a framework for understanding what all these approaches have in common. Say that an agent *believes U with respect to plausibility measure* Pl if $\mathrm{Pl}(U) > \mathrm{Pl}(\overline{U})$; that is, the agent believes U if U is more plausible than not. It is easy to see that, in general, this definition is not closed under conjunction. In the case of probability, for example, this definition just says that U is believed if the probability of U is greater than $1/2$. What condition on a plausibility measure Pl is needed to guarantee that this definition of belief is closed under conjunction? Trivially, the following restriction does the trick:

Pl4″. If $\mathrm{Pl}(U_1) > \mathrm{Pl}(\overline{U_1})$ and $\mathrm{Pl}(U_2) > \mathrm{Pl}(\overline{U_2})$, then $\mathrm{Pl}(U_1 \cap U_2) > \mathrm{Pl}(\overline{U_1 \cap U_2})$.

I actually want a stronger version of this property, to deal with *conditional* beliefs. An agent believes *U conditional on V*, if given V, U is more plausible than $\overline{U}$, that is, if $\mathrm{Pl}(U|V) > \mathrm{Pl}(\overline{U}|V)$. In the presence of CPl5 (which I implicitly assume for this section), conditional beliefs are closed under conjunction if the following holds:

Pl4′. If $\mathrm{Pl}(U_1 \cap V) > \mathrm{Pl}(\overline{U_1} \cap V)$ and $\mathrm{Pl}(U_2 \cap V) > \mathrm{Pl}(\overline{U_2} \cap V)$, then $\mathrm{Pl}(U_1 \cap U_2 \cap V) > \mathrm{Pl}(\overline{U_1 \cap U_2} \cap V)$.

A more elegant requirement is the following:

Pl4. If U_1, U_2, and U_3 are pairwise disjoint sets, $\mathrm{Pl}(U_1 \cup U_2) > \mathrm{Pl}(U_3)$, and $\mathrm{Pl}(U_1 \cup U_3) > \mathrm{Pl}(U_2)$, then $\mathrm{Pl}(U_1) > \mathrm{Pl}(U_2 \cup U_3)$.

In words, Pl4 says that if $U_1 \cup U_2$ is more plausible than U_3 and if $U_1 \cup U_3$ is more plausible than U_2, then U_1 by itself is already more plausible than $U_2 \cup U_3$.

Remarkably, in the presence of Pl1–3, Pl4 and Pl4′ are equivalent:

Proposition 4.1: *([Friedman and Halpern, 1996a])* Pl *satisfies* ***P****l1–4 iff* Pl *satisfies Pl1–3 and Pl4′.*

Thus, for plausibility measures, Pl4 is necessary and sufficient to guarantee that conditional beliefs are closed under conjunction. Proposition 4.1 helps explain why all the notions of belief discussed above are closed under conjunction. More precisely, for each notion of belief discussed earlier, it is trivial to construct a plausibility measure Pl satisfying Pl4 that captures it: Pl give plausibility 1 to the events that are believed and plausibility 0 to the rest.

Pl4 is required for beliefs to be closed under finite intersection (i.e., finite conjunction). It does not guarantee closure under infinite intersection. This is a feature: beliefs are not always closed under infinite intersection. The classic example is the *lottery paradox* [Kyburg, 1961]: Consider a situation with infinitely many individuals, each of whom holds a ticket to a lottery. It seems reasonable to believe that individual i will not win, for any i, yet that someone will win. If E_i is the event that individual i does not win, this amounts to believing $E_1, E_2, E_3, \ldots$ and also believing $\cup_i \overline{E_i}$ (and not believing $\cap_i E_i$). It is easy to capture this with a plausibility measure. Let $W = \{w_1, w_2, \ldots\}$, where w_i is the world where individual i wins (so that $E_i = W - \{w_i\}$). Let Pl_{lot} be a plausibility measure that assigns plausibility 0 to the empty set, plausibility $1/2$ to all finite sets, and plausibility 1 to all infinite sets. It is easy to see that Pl_{lot} ratifies Pl4. Nevertheless, each of E_1 is believed according to Pl_{lot}, as is $\cup_i \overline{E_i}$.

As shown in [Friedman *et al.*, 2000], the key property that guarantees that (conditional) beliefs are closed under infinite intersection is the following generalization of Pl4:

Pl4*. For any index set I such that $0 \in I$, if $\{U_i : i \in I\}$ are pairwise disjoint sets, $U = \cup_{i \in I} U_i$, and for all $i \in I - \{0\}$, $\mathrm{Pl}(U - U_i) > \mathrm{Pl}(U_i)$, then $\mathrm{Pl}(U_0) > \mathrm{Pl}(U - U_0)$.

Because Pl4* does not hold for Pl_{lot}, it can be used to represent the lottery paradox. Because Pl4* does hold for the plausibility measure corresponding to beliefs in Kripke structure, belief in Kripke structures is closed under infinite conjunction. A countable version of Pl4* holds for σ-additive probability measures, which is why probability-1 beliefs are closed under countable conjunctions (but not necessarily under arbitrary infinite conjunctions).

4.2 Belief Revision

An agent's beliefs change over time. Conditioning has been the standard approach to modeling this change in the context of probability. However, conditioning has been argued to be inapplicable when it comes to belief revision, because an agent may learn something inconsistent with her beliefs. This would amount to conditioning on a set of measure 0. As a consequence, finding appropriate models of belief change has been an active area in philosophy and in both artificial intelligence [Gärdenfors, 1988; Katsuno and Mendelzon, 1991]. In the literature, two models have been studied in detail: *Belief revision* [Alchourrón *et al.*, 1985; Gärdenfors, 1988] attempts to describe how an agent should accommodate a new belief (possibly inconsistent with his other beliefs) about a static world. *Belief update* [Katsuno and Mendelzon, 1991], on the other hand, attempts to describe how an agent should change his beliefs as a result of learning about a change in the world.

Belief revision and belief update describe only two of the many ways in which beliefs can change. Using plausibility, it is possible to construct a general framework for reasoning about belief change (see [Friedman and Halpern, 1997]). The key point is that it is possible to describe belief changing using conditioning with plausibility, even though it cannot be done with probability. Starting with a conditional plausibility measure satisfying Pl4 (this is necessary for belief to have the right properties) and conditioning on new information gives us a general model of belief change. Belief revision and belief update can be captured by putting appropriate constraints on the initial plausibility [Friedman and Halpern, 1999]. The same framework can be used to capture other notions of belief change, such as a general *Markovian* models of belief change [Friedman and Halpern, 1996b] and belief change with unreliable observations [Boutilier *et al.*, 1998]. The key point is that belief change simply becomes conditioning (and iterated belief change becomes iterated conditioning).

4.3 Default Reasoning

It has been argued that *default reasoning* plays a major role in commonsense reasoning. Perhaps not surprisingly, there have been many approaches to default reasoning proposed in the literature (see [Gabbay *et al.*, 1993; Ginsberg, 1987]). Many of the recent approaches to giving semantics to defaults can be viewed as considering structures of the form (W, X, π), where W is a set of possible worlds, $\pi(w)$ is a truth assignment to primitive propositions for each world $w \in W$, and X can be viewed as a "measure" on W. Some examples of X include possibility measures [Dubois and Prade, 1991], κ-rankings [Goldszmidt and Pearl, 1996], *parameterized probability distributions* [Pearl, 1989] (these are sequences of probability distributions; the resulting approach is more commonly known as ϵ-*semantics*), and *preference orders* [Kraus *et al.*, 1990; Lewis, 1973].

Somewhat surprisingly, all of these approaches are characterized by the six axioms and inference rules, which have been called the *KLM properties* (since they were discussed by Kraus, Lehmann, and Magidor 1990). Assume (as is typical in the literature) that defaults are expressed in terms of an operator $\rightarrow$, where $\varphi \rightarrow \psi$ is read "if φ then typically/likely/by default ψ. For example, the default "birds typically fly" is represented *Bird*$\rightarrow$*Fly*. We further assume for now that the formulas φ and ψ that appear in defaults come from some propositional language $\mathcal{L}$ with a consequence relation $\vdash_{\mathcal{L}}$.

LLE. If $\vdash_{\mathcal{L}} \varphi \Leftrightarrow \varphi'$, then from $\varphi \rightarrow \psi$ infer $\varphi' \rightarrow \psi$ (left logical equivalence).

RW. If $\vdash_{\mathcal{L}} \psi \Rightarrow \psi'$, then from $\varphi \rightarrow \psi$ infer $\varphi \rightarrow \psi'$ (right weakening).

REF. $\varphi \rightarrow \varphi$ (reflexivity).

AND. From $\varphi \rightarrow \psi_1$ and $\varphi \rightarrow \psi_2$ infer $\varphi \rightarrow \psi_1 \wedge \psi_2$.

OR. From $\varphi_1 \rightarrow \psi$ and $\varphi_2 \rightarrow \psi$ infer $\varphi_1 \vee \varphi_2 \rightarrow \psi$.

CM. From $\varphi \rightarrow \psi_1$ and $\varphi \rightarrow \psi_2$ infer $\varphi \wedge \psi_2 \rightarrow \psi_1$ (cautious monotonicity).

LLE states that the syntactic form of the antecedent is irrelevant. Thus, if φ_1 and φ_2 are equivalent, we can deduce $\varphi_2 \rightarrow \psi$ from $\varphi_1 \rightarrow \psi$. RW describes a similar property of the consequent: If ψ (logically) entails ψ', then we can deduce $\varphi \rightarrow \psi'$ from $\varphi \rightarrow \psi$. This allows us to combine default and logical reasoning. REF states that φ is always a default conclusion of φ. AND states that we can combine two default conclusions. If we can conclude by default both ψ_1 and ψ_2 from φ, then we can also conclude $\psi_1 \wedge \psi_2$ from φ. OR states that we are allowed to reason by cases. If the same default conclusion follows from each of two antecedents, then it also follows from their disjunction. CM states that if ψ_1 and ψ_2 are two default conclusions of φ, then discovering that ψ_2 holds when φ holds (as would be expected, given the default) should not cause us to retract the default conclusion ψ_1.

The fact that the KLM properties characterize so many different semantic approaches has been viewed as rather surprising, since these approaches seem to capture quite different intuitions. As Pearl 1989 said of the equivalence between ϵ-semantics and preferential structures, "It is remarkable that two totally different interpretations of defaults yield identical sets of conclusions and identical sets of reasoning machinery." Plausibility measures help us understand why this should be so. In fact, plausibility measures provide a much deeper understanding of exactly what properties a semantic approach must have in order to be characterized by the KLM properties.

The first step to obtaining this understanding is to give semantics to defaults using plausibility. A *plausibility structure* is a tuple (W, Pl, π), where Pl is a plausibility measure on W. A conditional $\varphi \rightarrow \psi$ holds in this structure if either $\mathrm{Pl}([\![\varphi]\!]) = \bot$ or $\mathrm{Pl}([\![\varphi \wedge \psi]\!]) > \mathrm{Pl}([\![\varphi \wedge \neg\psi]\!])$ (where $[\![\sigma]\!]$ is the set of worlds satisfying the formula σ). This approach is just a generalization of the approach first used to define defaults with possibility measures [Dubois and Prade, 1991]. Note that if Pl satisfies CPl5, this is equivalent to saying that $\mathrm{Pl}([\![\psi]\!] | [\![\varphi]\!]) > \mathrm{Pl}([\![\neg\psi]\!] | [\![\varphi]\!])$ if $[\![\varphi]\!] \neq \bot$ (the implicit assumption here is that $[\![\varphi]\!] \in \mathcal{F}'$ iff $[\![\varphi]\!] \neq \bot$).

While this definition of defaults in terms of plausibility is easily seen to satisfy REF, RW, and LLE, it does not satisfy AND, OR, or CM in general. It is easy to construct counterexamples taking Pl to be a probability measure Pr (in which case the definition boils down to $\varphi \rightarrow \psi$ if $\Pr([\![\varphi]\!]) = 0$ or $\Pr([\![\psi]\!] | [\![\varphi]\!]) > 1/2$). As observed earlier, if Pl satisfies Pl4 (which it does not in general if Pl is a probability measure), then the AND rule is satisfied. As shown in [Friedman and Halpern, 1996a], Pl4 also suffices to guarantee CM (cautious monotonicity). The only additional property that is needed to guarantee that OR holds is the following:

Pl5. If $\mathrm{Pl}(U) = \mathrm{Pl}(V) = \bot$, then $\mathrm{Pl}(U \cup V) = \bot$.

A plausibility structure (W, Pl, π) is *qualitative* if Pl satisfies Pl1–5. In [Friedman and Halpern, 1996a], it is shown that a necessary and sufficient condition for a collection of plausibility structures to satisfy the KLM properties is that they be qualitative. More precisely, given a class $\mathcal{P}$ of plausibility structures, a default d is *entailed by a set Δ of defaults in $\mathcal{P}$*, written $\Delta \models_{\mathcal{P}} d$, if all structures in $\mathcal{P}$ that satisfy all the defaults in Δ also satisfy d. Let $\mathcal{S}^{QPL}$ consist of all qualitative plausibility structures. Write $\Delta \vdash_{\mathbf{P}} \varphi \rightarrow \psi$ if $\varphi \rightarrow \psi$ is provable from Δ using the KLM properties.

Theorem 4.2: [Friedman and Halpern, 1996a] *$\mathcal{S} \subseteq \mathcal{S}^{QPL}$ if and only if for all Δ, φ, and ψ, if $\Delta \vdash_{\mathbf{P}} \varphi \rightarrow \psi$ then $\Delta \models_{\mathcal{S}} \varphi \rightarrow \psi$.*

In [Friedman and Halpern, 1996a], it is shown that possibility structures, κ-structures, preferential structures, and PPDs can all be viewed as qualitative plausibility structures. Theorem 4.2 thus shows why the KLM properties hold in all these cases. Why are there no further properties (that is, why are the KLM properties not only sound, but complete)? To show that the KLM properties are complete with respect to a class $\mathcal{S}$ of structures, we have to ensure that $\mathcal{S}$ contains "enough" structures. In particular, if $\Delta \not\vdash_{\mathbf{P}} \varphi \rightarrow \psi$, we want to ensure that there is a plausibility structure $PL \in \mathcal{S}$ such that $PL \models_{PL} \Delta$ and $PL \not\models_{PL} \varphi \rightarrow \psi$. The following weak condition on $\mathcal{S}$ guarantees this.

Definition 4.3: We say that $\mathcal{S}$ is *rich* if for every collection $\varphi_1, \ldots, \varphi_n$, $n > 1$, of mutually exclusive formulas, there is a plausibility structure $PL = (W, \mathrm{Pl}, \pi) \in \mathcal{S}$ such that:

$$\mathrm{Pl}([\![\varphi_1]\!]) > \mathrm{Pl}([\![\varphi_2]\!]) > \cdots > \mathrm{Pl}([\![\varphi_n]\!]) = \bot. \blacksquare$$

The richness condition is quite mild. Roughly speaking, it says that we do not have *a priori* constraints on the relative plausibilities of a collection of disjoint sets. It is easily seen to hold for the plausibility structures that arise from preferential structures (resp., possibility structures, κ-structures, PPDs). More importantly, richness is a necessary and sufficient condition to ensure that the KLM properties are complete.

Theorem 4.4: [Friedman and Halpern, 1996a] *A set $\mathcal{S}$ of qualitative plausibility structures is rich if and only if for all finite Δ and defaults $\varphi \rightarrow \psi$, we have that $\Delta \models_{\mathcal{S}} \varphi \rightarrow \psi$ implies $\Delta \vdash_{\mathbf{P}} \varphi \rightarrow \psi$.*

This result shows that if the KLM properties are sound with respect to a class of structures, then they are almost inevitably complete as well. More generally, Theorems 4.2 and 4.4 explain why the KLM properties are sound and complete for so many approaches.

The discussion up to now has focused on propositional defaults, but using plausibility, it is fairly straightforward to extend to the first-order case; see [Friedman *et al.*, 2000].

5 Expectation and Decision Theory

Agents must make decisions. Perhaps the best-known rule for decision making is that of maximizing expected utility. This requires that agents have probabilities for many events of interest, and numerical utilities. But many other decision rules have been proposed, including minimax, regret minimization, and rules that involve representations of uncertainty other than probability. Again, using plausibility allows us to understand what is required to get various desirable properties of decision rules.

Since expectation plays such a key role in maximizing expected utility, I start by considering expectation. Given a probability measure μ on some sample space W, the corresponding expectation function E_μ maps gambles over W

(that is, random variables with domain W and range the reals) to reals. There are a number of equivalent definitions of E_μ. The standard one is

$$E_\mu(X) = \sum_{x \in \mathcal{V}(X)} x\mu(X = x). \quad (1)$$

(Here I am implicitly assuming that $X = x$ (that is, the set $\{w : X(w) = x\}$) is measurable.)

As is well known, E_μ is linear ($E_\mu(aX+Y) = aE_\mu(X)+E_\mu(Y)$), monotonic (if $X \leq Y$, then $E_\mu(X) \leq E_\mu(Y)$), and it maps a constant function to its value (that is, $\tilde{a}$ is the gamble that maps all elements of W to a, the $E_\mu(\tilde{a}) = a$). Moreover, these three properties characterize probabilistic expectation functions. If an expectation function E has these properties, then $E = E_\mu$ for some probability measure μ.

A *W–D' expectation function* is simply a mapping from random variables with domain W and range some ordered set D' to D'. Here I focus on expectation functions that are generated by some plausibility measure, just as E_μ is generated from μ, using a definition in the spirit of (1). To do this, we need analogues of $+$ and $\times$, much in the spirit of (but not identical to) the $\oplus$ and $\otimes$ used in the definition of algebraic cps.

Definition 5.1 : An *expectation domain* is a tuple $(D, D', \boxplus, \boxtimes)$, where D and D' are sets ordered by $\leq_D$ and $\leq_{D'}$, respectively, D is a set of plausibility values (so that it has elements $\bot$ and $\top$ such that $\bot \leq_D d \leq_D \top$ for all $d \in D$), $\boxplus : D' \times D' \to D'$ and $\boxtimes : D' \times D \to D'$. ∎

Given an expectation domain $(D, D', \boxplus, \boxtimes)$ and a plausibility measure Pl on some set W, it is possible to define a W–D' expectation function E_{Pl} by using the obvious analogue of (1), replacing $+$ by $\boxplus$ and $\times$ by $\boxtimes$.

What does this buy us? For one thing, we can try to characterize the properties of $\boxplus$ and $\boxtimes$ that give E_{Pl} properties of interest, such as linearity and monotonicity (see [Halpern, 2000b] for details). For another, it turns out that all standard decision rules can be expressed as expected utility maximization with respect to an appropriate choice of plausibility measure, $\boxplus$, and $\boxtimes$. To make this precise, assume that there is a set $\mathcal{A}$ of possible actions that an agent can perform. An action $\mathbf{a}$ maps a world $w \in W$ to an outcome. For simplicity, I identify outcomes with world-action pairs $(w, \mathbf{a})$. Assume that the agent has a utility function u on outcomes. In the examples below, the range of the utility function is the reals but, in general, it can be an arbitrary partially ordered set D'. Let $u_{\mathbf{a}}$ be the random variable such that $u_{\mathbf{a}}(w) = u(w, \mathbf{a})$. The agent is uncertain about the actual world; this uncertainty is represented by some plausibility measure. The question is which action the agent should choose.

As I said earlier, if the agent's uncertainty is represented by a probability measure μ, the standard decision rule is to choose the action that maximizes expected utility. That is, we choose an action $\mathbf{a}$ such that $E_\mu(\mathbf{a}) \geq E_\mu(\mathbf{a}')$ for all $\mathbf{a}' \in \mathcal{A}$. However, there are other well-known decision rules.

- For minimax, let *worst*$(\mathbf{a}) = \min\{u_{\mathbf{a}}(w) : w \in W\}$; *worst*$(\mathbf{a})$ is the utility of the worst-case outcome if $\mathbf{a}$ is performed. This too leads to a total preference order on actions, where $\mathbf{a}$ is preferred to $\mathbf{a}'$ if *worst*$(\mathbf{a}) \geq$ *worst*$(\mathbf{a}')$. The minimax rule says to choose the action $\mathbf{a}$ (or one of them, in case of ties) such that *worst*$(\mathbf{a})$ is highest. The action chosen according to this rule is the one with the best worst-case outcome. Notice that minimax makes sense no matter how uncertainty is represented. Now take $D = \{0, 1\}$ and $D' = I\!R$, both with the standard order, and consider the plausibility measure Pl_{mm}, where $\mathrm{Pl}_{mm}(U) = 1$ if $U \neq \emptyset$ and $\mathrm{Pl}_{mm}(\emptyset) = 0$. Let $\boxplus$ be min and let $\boxtimes$ be multiplication. With this choice of $\boxplus$ and $\boxtimes$, it is easy to see that $E_{\mathrm{Pl}_{mm}}(u_{\mathbf{a}}) =$ *worst*$(\mathbf{a})$, so expected utility maximization with respect to Pl_{mm} is minimax.
- As a first step to defining regret minimization, for each world w, let $\mathbf{a}_w$ be an action that gives the best outcome in world w; that is, $u(w, \mathbf{a}_w) \geq u(w, \mathbf{a})$ for all $\mathbf{a} \in \mathcal{A}$. The *regret* of $\mathbf{a}$ in world w is $u(w, \mathbf{a}_w) - u(w, \mathbf{a})$; that is, the regret of $\mathbf{a}$ in w is the difference between the utility of performing the best action in w (the action that the agent would perform, presumably, if she knew the actual world was w) and that of performing $\mathbf{a}$ in w. Finally, define *regret*$(\mathbf{a}) = \max_{w \in W}$ *regret*$(\mathbf{a}, w)$. Intuitively, if *regret*$(\mathbf{a}) = k$, then $\mathbf{a}$ is guaranteed to be within k of the best action the agent could perform, even if she she knew exactly what the actual world was. The decision rule of minimizing regret chooses the action $\mathbf{a}$ such that *regret*$(\mathbf{a})$ is a minimum.

  To express regret in terms of maximizing expected utility, it is easiest to assume that for each action $\mathbf{a}$, $\max_{w \in W} u_{\mathbf{a}}(w) = 1$. This assumption is without loss of generality: if $u'(w, \mathbf{a}) = u(w, \mathbf{a}) - \max_{w' \in W} u(w', \mathbf{a}) + 1$, then $\max_{w \in W} u'_{\mathbf{a}}(w) = 1$, and minimizing regret with respect u' gives the same result as minimizing regret with respect to u. With this assumption, take $D = [-\infty, 1]$ with the standard ordering and $D' = I\!R$ with the reverse ordering, that is $x <_{D'} y$ if $x > y$. Let $\mathrm{Pl}_{reg}(U) = \max_{w \in U, \mathbf{a} \in \mathcal{A}} u(\mathbf{a}, w)$, let $a \boxtimes b = a - b$, and let $a \boxplus b = \min(a, b)$. Intuitively, $\mathrm{Pl}(U) \boxtimes b$ is the regret an agent would feel if she is given utility b but could have performed the action that would give her the best outcome on her choice of world in U. With this choice of $\boxplus$ and $\boxtimes$, it is easy to see that $E_{\mathrm{Pl}_{reg}}(u_{\mathbf{a}}) =$ *regret*$(\mathbf{a})$, so expected utility maximization with respect to Pl_{reg} is just regret minimization (given the ordering on D').
- Suppose that uncertainty is represented by a set $\mathcal{P}$ of probability measures indexed by some set I. There are two natural ways to get a partial order on actions from $\mathcal{P}$ and a real-valued utility u. Define $\succeq^1_{\mathcal{P}}$ so that $\mathbf{a} \succeq^1_{\mathcal{P}} \mathbf{a}'$ iff $\min_{\mu \in \mathcal{P}} E_\mu(u_{\mathbf{a}}) \geq \max_{\mu \in \mathcal{P}} E_\mu(u_{s_{\mathbf{a}}})$. That is, $\mathbf{a}$ is preferred to $\mathbf{a}'$ if the expected utility of performing $\mathbf{a}$ is at least that of performing $\mathbf{a}'$, regardless which probability measure in $\mathcal{P}$ describes the actual probability. Naturally, $\succeq^1_{\mathcal{P}}$ is only a partial order on actions. A more refined partial order can be obtained as follows: Define $\mathbf{a} \succeq^2_{\mathcal{P}} \mathbf{a}'$ if $E_\mu(u_{\mathbf{a}}) \geq E_\mu(u_{\mathbf{a}'})$ for all $\mu \in \mathcal{P}$. It is easy to show that if $\mathbf{a} \succeq^1_{\mathcal{P}} \mathbf{a}'$ then $\mathbf{a} \succeq^2_{\mathcal{P}} \mathbf{a}'$, although the converse may not hold. For example, suppose that $\mathcal{P} = \{\mu, \mu'\}$ and actions $\mathbf{a}$ and $\mathbf{a}'$ are such

that $E_\mu(u_{\mathbf{a}}) = 2$, $E_{\mu'}(u_{\mathbf{a}}) = 4$, $E_\mu(u_{\mathbf{a}'}) = 1$, and $E_{\mu'}(u_{\mathbf{a}'}) = 3$. Then $\mathbf{a}$ and $\mathbf{a}'$ are incomparable according to $\succeq^1_{\mathcal{P}}$, but $\mathbf{a} \succeq^2_{\mathcal{P}} \mathbf{a}'$.

Let the set D of plausibility values be that used for $\mathrm{Pl}_{\mathcal{P}}$, that is, the functions from I to $[0,1]$, with the pointwise ordering. Let D' be the functions from I to $\mathbb{R}$, let $\boxplus$ be pointwise addition, and let $\boxtimes$ be pointwise multiplication. The difference between $\succeq^1_{\mathcal{P}}$ and $\succeq^2_{\mathcal{P}}$ is captured by considering two different orders on D'. For $\succeq^1_{\mathcal{P}}$, order D' by $>^1_{D'}$, where $f \geq^1_{D'} g$ if $\min_{i \in I} f(i) \geq \max_{i \in I} g(i)$, while for $\succeq^2_{\mathcal{P}}$, order D' by $>^2_{D'}$, where $f \geq^2_{D'} g$ if $f(i) \geq g(i)$ for all $i \in I$. If $E_{\mathrm{Pl}_{\mathcal{P}}}$ is the expectation function corresponding to this definition of $\boxplus$ and $\boxtimes$, then it is easy to see that $E_{\mathrm{Pl}_{\mathcal{P}}}(u_{\mathbf{a}}) \geq^j_{D'} E_{\mathrm{Pl}_{\mathcal{P}}}(u_{\mathbf{a}'})$ iff $\mathbf{a} \succeq^j_{\mathcal{P}} \mathbf{a}'$, for $j = 1, 2$.

It can be shown that every partial order on actions can be represented as the ordering induced by expected utility according to some plausibility measure on W. That is, given some partial order $\succeq$ on actions that can be taken in some set W of possible worlds, there is a plausibility measure Pl on W and expectation domain $(D, D', \boxplus, \boxtimes)$ such that the range of Pl is D and a utility function on $W \times \mathcal{A}$ with range D' such that $E_{\mathrm{Pl}}(u_{\mathbf{a}}) \geq E_{\mathrm{Pl}}(u_{\mathbf{a}'})$ iff $\mathbf{a} \succeq \mathbf{a}'$. Thus, viewing decision rules as instances of expected utility maximization with respect to the appropriate expectation function provides a general framework in which to study decision rules. For example, it becomes possible to ask what properties of an expectation domain are needed to get various of Savage's 1954 postulates. I hope to report on this in future work.

6 Compact Representation of Uncertainty

Suppose that W is a set of possible worlds characterized by n binary random variables $\mathcal{X} = \{X_1, \ldots, X_n\}$ (or, equivalently, n primitive propositions). That is, a world $w \in W$ is a tuple $(x_1, \ldots, x_n)$, where $x_i \in \{0,1\}$ is the value of X_i. That means that there are 2^n worlds in W, say $w_1, \ldots, w_{2^n}$. A naive description of a probability measure on W requires $2^n - 1$ numbers, $\alpha_1, \ldots, \alpha_{2^n-1}$, where α_i is the probability of world w_i. (Of course, the probability of w_{2^n} is determined by the other probabilities, since they must sum to 1.)

If n is relatively small, describing a probability measure in this naive way is not so unreasonable, but if n is, say, 1000 (certainly not unlikely in many practical applications), then it is completely infeasible. One of the most significant recent advances in AI has been in work on *Bayesian networks* [Pearl, 1988], a tool that allows probability measures to be represented in a compact way and manipulated in a computationally feasible way. I briefly review Bayesian networks here and then discuss the extent to which the ideas can be applied to other representations of uncertainty. More details can be found in [Halpern, 2000a].

Recall that a (qualitative) *Bayesian network* (sometimes called a *belief network*) is a *dag*, that is, a directed acyclic graph, whose nodes are labeled by random variables. Informally, the edges in a Bayesian network can be thought of as representing causal influence.

Given a Bayesian network G and a node X in G, think of the *ancestors* of X in the graph as those random variables that have a potential influence on X. This influence is mediated through the *parents* of X, those ancestors of X directly connected to X. That means that X should be conditionally independent of its ancestors, given its parents. The formal definition requires, in fact, that X be independent not only of its ancestors, but of its *nondescendants*, given its parents, where the nondescendants of X are those nodes Y such that X is not the ancestor of Y.

Definition 6.1: Given a qualitative Bayesian network G, let $\mathrm{Par}_G(X)$ be the parents of the random variable X in G; let $G_{\mathrm{Des}}(X)$ be all the *descendants* of X, that is, X and all those nodes Y such that X is an ancestor of Y; let $\mathrm{ND}_G(X)$, the nondescendants of X in G, consist of $\mathcal{X} - \mathrm{Des}_G(X)$. The Bayesian network G *(qualitatively) represents* the probability measure μ if X is conditionally independent of its nondescendants given its parents, for all $X \in \mathcal{X}$. ∎

A qualitative Bayesian network G gives qualitative information about dependence and independence, but does not actually give the values of the conditional probabilities. A quantitative Bayesian network provides more quantitative information, by associating with each node X in G a *conditional probability table (cpt)* that quantifies the effects of the parents of X on X. For example, if X's parents in G are Y and Z, then the cpt for X would have an entry denoted $d_{Y=j,Z=k}$ for all $(j,k) \in \{0,1\}^2$. As the notation is meant to suggest, $d_{Y=j \cap Z=k} = \mu(X = 1 | Y = j \cap Z = k)$ for the plausibility measure μ represented by G. (Of course, there is no need to have an entry for $\mu(X = 0 | Y = j \cap Z = k)$, since this is just $1 - \mu(X = 1 | Y = j \cap Z = k)$.) Formally, a *quantitative Bayesian network* is a pair (G, f) consisting of a qualitative Bayesian network G and a function f that associates with each node X in G a cpt, where there is an entry in the interval $[0,1]$ in the cpt for each possible setting of the parents of X. If X is a root of G, then the cpt for X can be thought of as giving the unconditional probability that $X = 1$.

Definition 6.2 : A quantitative Bayesian network (G, f) *(quantitatively) represents*, or *is compatible with*, the probability measure μ if G qualitatively represents μ and the cpts agree with μ in that, for each random variable X, the entry in the cpt for X given some setting $Y_1 = y_1, \ldots, Y_k = y_k$ of its parents is $\mu(X = 1 | Y_1 = y_1 \cap \ldots \cap Y_k = y_k)$ if $\mu(Y_1 = y_1 \cap \ldots \cap Y_k = y_k) \neq 0$. (It does not matter what the cpt entry for $Y_1 = y_1, \ldots, Y_k = y_k$ is if $\mu(Y_1 = y_1 \cap \ldots \cap Y_k = y_k) = 0$.) ∎

It can easily be shown using the chain rule for probability (see, for example, [Pearl, 1988]) that if (G, f) quantitatively represents μ, then μ can be completely reconstructed from (G, f). More precisely, the 2^n values $\mu(X_1 = x_1 \cap \ldots \cap X_n = x_n)$ can be computed from (G, f); from these values, $\mu(U)$ can be computed for all $U \subseteq W$.

Bayesian networks for probability have a number of important properties:

1. Every probability measure is represented by a qualitative Bayesian network (in fact, in general there are many qualitative Bayesian networks that represent a given probability measure).

2. A qualitative Bayesian network that represents a probability measure μ can be extended to a quantitative Bayesian network that represents μ, by adding cpts.

3. A quantitative Bayesian network represents a unique probability measure. This is important because if a world is characterized by the values of n random variables, so that there are 2^n worlds, a quantitative Bayesian network can often represent a probability measure using far fewer than 2^n numbers. If a node in the network has k parents, then its conditional probability table has 2^k entries. Therefore, if each node has at most k parents in the graph, then there are at most $n2^k$ entries in all the cpts. If k is small, then $n2^k$ can be much smaller than $2^n - 1$.

4. A Bayesian network supports efficient algorithms for computing conditional probabilities of the form $\Pr(X_i = x_i | X_j = x_j)$; that is, they allow for efficient evaluation of probabilities given some information.

To what extent is probability necessary to achieve these properties? More precisely, what properties of probability are needed to achieve them? Here again, plausibility measures allow us to answer this question.

Given a cps $(W, \mathcal{F}, \mathcal{F}', \mathrm{Pl})$, $U, V \in \mathcal{F}$ are *plausibilistically independent given V' (with respect to* Pl*)*, written $I_{\mathrm{Pl}}(U, V | V')$, if $V \cap V' \in \mathcal{F}'$ implies $\mathrm{Pl}(U | V \cap V') = \mathrm{Pl}(U | V')$ and $U \cap V' \in \mathcal{F}'$ implies $\mathrm{Pl}(V | U \cap V') = \mathrm{Pl}(V | V')$. This definition is meant to capture the intuition that (conditional on V') U and V are independent if learning about U gives no information about V and learning about V gives no information about U. Note the explicitly symmetric nature of the definition. In the case of probability, if learning about U gives no information about V, then it is immediate that learning about V gives no information about U. This does not hold for an arbitrary plausibility measure.[3]

If $\mathbf{X} = \{X_1, \ldots, X_n\}$, $\mathbf{Y} = \{Y_1, \ldots, Y_m\}$, and $\mathbf{Z} = \{Z_1, \ldots, Z_k\}$ are sets of random variables, then $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent given $\mathbf{Z}$ (with respect to Pl) if $X_1 = x_1 \cap \ldots \cap X_n = x_n$ is conditionally independent of $Y_1 = y_1 \cap \ldots \cap Y_m = y_m$ given $Z_1 = z_1 \cap \ldots \cap Z_k = z_k$ for all choices of $x_1, \ldots, x_n, y_1, \ldots, y_m, z_1, \ldots, z_k$.

With these definitions, the notion of a qualitative Bayesian network as defined in Definition 6.1 makes perfect sense if the probability measure μ is replaced by a plausibility measure Pl everywhere. The following result shows that representation by a qualitative Bayesian network is possible not just in the case of probability, but for any algebraic cps.

[3] An equivalent definition of U and V being independent with respect to a probability measure μ is that $\mu(U \cap V | V') = \mu(U | V') \times \mu(V | V')$. However, I want to give a definition of independence that does not require an analogue to multiplication. But even in an algebraic cps, the requirement that $\mu(U \cap V | V') = \mu(U | V') \otimes \mu(V | V')$ is not always equivalent to the definition given here (see [Halpern, 2000a]). Also note that if $V \cap V' \notin \mathcal{F}'$ (in the case of probability, this would correspond to $V \cap V'$ having probability 0), then $\mathrm{Pl}(U | V \cap V')$ is not defined. In this case, there is no requirement that $\mathrm{Pl}(U | V \cap V') = \mathrm{Pl}(U | V')$. A similar observation holds if $U \cap V' \notin \mathcal{F}'$.

Theorem 6.3: *([Halpern, 2000a]) If $(W, \mathcal{F}, \mathcal{F}', \mathrm{Pl})$ is an algebraic cps, then there is a qualitative Bayesian network that represents* Pl.

Clearly a qualitative Bayesian network that represents Pl can be extended to a quantitative Bayesian network (G, f) that represents Pl by filling in the conditional plausibility tables. But does a quantitative Bayesian network (G, f) represents a unique (algebraic) plausibility measure? Recall that, for the purposes of this section, I have taken W to consist of the 2^n worlds characterized by the n binary random variables in $\mathcal{X}$. Let $PL_{D,\otimes,\oplus}$ consist of all algebraic standard cps's of the form $(W, \mathcal{F}, \mathcal{F}', \mathrm{Pl})$, where $\mathcal{F} = 2^W$, so that all subsets of W are measurable, and the range of Pl is D. With this notation, the question becomes whether a quantitative Bayesian network (G, f) such that the entries in the cpts are in D determines a unique element in $PL_{D,\oplus,\otimes}$. It turns out that the answer is yes, provided that $(D, \oplus, \otimes)$ satisfies some conditions. The conditions are similar in spirit to Alg1–4, except that now they are conditions on $(D, \oplus, \otimes)$, rather than conditions on a plausibility measure; I omit the details here (again, see [Halpern, 2000a]). The key point is that these conditions are sufficient to allow an arbitrary plausibility measure to have a compact representation. Moreover, since the typical algorithms in probabilistic Bayesian networks use only algebraic properties of $+$ and $\times$, they apply with essentially no change to algebraic plausibility measures.

7 Conclusions

There is no reason to believe that one representation of uncertainty is best for all applications. This makes it useful to have a framework in which to compare representations. As I hope I have convinced the reader, plausibility measures give us such a framework, and provide a vantage point from which to look at representations of uncertainty and understand what makes them tick—what properties of each one are being used to get results of interest. More discussion of these and related topics can be found in [Halpern, 2000c].

References

[Adams, 1966] E. Adams. Probability and the logic of conditionals. In J. Hintikka and P. Suppes, editors, *Aspects of Inductive Logic*, pages 265–316. North Holland, 1966.

[Alchourrón *et al.*, 1985] C. E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet functions for contraction and revision. *Journal of Symbolic Logic*, 50:510–530, 1985.

[Blume *et al.*, 1991] L. Blume, A. Brandenburger, and E. Dekel. Lexicographic probabilities and choice under uncertainty. *Econometrica*, 59(1):61–79, 1991.

[Boutilier *et al.*, 1998] C. Boutilier, J. Y. Halpern, and N. Friedman. Belief revision with unreliable observations. In *Proc.AAAI '98*, pages 127–134, 1998.

[Brandenburger and Dekel, 1987] A. Brandenburger and E. Dekel. Common knowledge with probability 1. *Journal of Mathematical Economics*, 16:237–245, 1987.

[Brandenburger and Keisler, 2000] A. Brandenburger and J. Keisler. Epistemic conditions for iterated admissibility. Unpublished manuscript, 2000.

[Brandenburger, 1999] A. Brandenburger. On the existence of a "complete" belief model. Working Paper 99-056, Harvard Business School, 1999.

[Dubois and Prade, 1990] D. Dubois and H. Prade. An introduction to possibilistic and fuzzy logics. In G. Shafer and J. Pearl, editors, *Readings in Uncertain Reasoning*, pages 742–761. Morgan Kaufmann, 1990.

[Dubois and Prade, 1991] D. Dubois and H. Prade. Possibilistic logic, preferential models, non-monotonicity and related issues. In *Proc. IJCAI '91*, pages 419–424. 1991.

[Friedman and Halpern, 1995] N. Friedman and J. Y. Halpern. Plausibility measures: a user's guide. In *Proc. UAI '95*, pages 175–184. 1995.

[Friedman and Halpern, 1996a] N. Friedman and J. Y. Halpern. Plausibility measures and default reasoning. In *Proc. AAAI '96*, pages 1297–1304. 1996. To appear, *Journal of the ACM*. Full version available at http://www.cs.cornell.edu/home/halpern.

[Friedman and Halpern, 1996b] N. Friedman and J. Y. Halpern. A qualitative Markov assumption and its implications for belief change. In *Proc. UAI '96*, pages 263–273, 1996.

[Friedman and Halpern, 1997] N. Friedman and J. Y. Halpern. Modeling belief in dynamic systems. Part I: foundations. *Artificial Intelligence*, 95:257–316, 1997.

[Friedman and Halpern, 1999] N. Friedman and J. Y. Halpern. Modeling belief in dynamic systems. Part II: revision and update. *Journal of A.I. Research*, 10:117–167, 1999.

[Friedman *et al.*, 2000] N. Friedman, J. Y. Halpern, and D. Koller. First-order conditional logic for default reasoning revisited. *ACM Trans. on Computational Logic*, 1(2):175–207, 2000.

[Gabbay *et al.*, 1993] D. M. Gabbay, C. J. Hogger, and J. A. Robinson, editors. *Nonmonotonic Reasoning and Uncertain Reasoning*, volume 3 of *Handbook of Logic in Artificial Intelligence and Logic Programming*. Oxford University Press, 1993.

[Gärdenfors, 1988] P. Gärdenfors. *Knowledge in Flux*. MIT Press, 1988.

[Ginsberg, 1987] M. L. Ginsberg, editor. *Readings in Nonmonotonic Reasoning*. Morgan Kaufmann, 1987.

[Goldszmidt and Pearl, 1992] M. Goldszmidt and J. Pearl. Rank-based systems: A simple approach to belief revision, belief update and reasoning about evidence and actions. In *KR '92*, pages 661–672, 1992.

[Goldszmidt and Pearl, 1996] M. Goldszmidt and J. Pearl. Qualitative probabilities for default reasoning, belief revision, and causal modeling. *Artificial Intelligence*, 84:57–112, 1996.

[Halpern, 2000a] J. Y. Halpern. Conditional plausibility measures and Bayesian networks. In *Proc. UAI 2000*, pages 247–255. 2000.

[Halpern, 2000b] J. Y. Halpern. On expectation. Unpublished manuscript, 2000.

[Halpern, 2000c] J. Y. Halpern. Reasoning about uncertainty. Book manuscript, 2000.

[Halpern and Fagin, 1992] J. Y. Halpern and R. Fagin. Two views of belief: belief as generalized probability and belief as evidence. *Artificial Intelligence*, 54:275–317, 1992.

[Hintikka, 1962] J. Hintikka. *Knowledge and Belief*. Cornell University Press, 1962.

[Katsuno and Mendelzon, 1991] H. Katsuno and A. Mendelzon. On the difference between updating a knowledge base and revising it. In *Principles of Knowledge Representation and Reasoning: Proc. KR '91*, pages 387–394, 1991.

[Kraus *et al.*, 1990] S. Kraus, D. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Kyburg, 1961] H. E Kyburg, Jr. *Probability and the Logic of Rational Belief*. Wesleyan University Press, 1961.

[Lehmann and Magidor, 1992] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, 55:1–60, 1992.

[Lewis, 1973] D. K. Lewis. *Counterfactuals*. Harvard University Press, 1973.

[McGee, 1994] V. McGee. Learning the impossible. In E. Eells and B. Skyrms, editors, *Probability and Conditionals*. Cambridge University Press, 1994.

[Myerson, 1986] R. Myerson. Multistage games with communication. *Econometrica*, 54:323–358, 1986.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.

[Pearl, 1989] J. Pearl. Probabilistic semantics for nonmonotonic reasoning: a survey. In *Proc. KR '89*, pages 505–516, 1989.

[Popper, 1968] K. R. Popper. *The Logic of Scientific Discovery (revised edition)*. Hutchison, London, 1968. First appeared as *Logik der Forschung*, 1934.

[Savage, 1954] L. J. Savage. *Foundations of Statistics*. John Wiley & Sons, 1954.

[Shafer, 1976] G. Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[Spohn, 1988] W. Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. In W. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change, and Statistics*, volume 2, pages 105–134. Reidel, 1988.

[van Fraassen, 1995] B. C. van Fraassen. Fine-grained opinion, probability, and the logic of full belief. *Journal of Philosophical Logic*, 24:349–377, 1995.

Robust Translation of Spontaneous Speech: A Multi-Engine Approach

Wolfgang Wahlster
DFKI
Stuhlsatzenhausweg 3
D-66123 Saarbrücken, Germany
wahlster@dfki.de

Abstract

Verbmobil is a speaker-independent and bidirectional speech-to-speech translation system for spontaneous dialogs that can be accessed via GSM mobile phones. It handles dialogs in three business-oriented domains, with context-sensitive translation between four languages (English, German, Japanese, and Chinese). We show that in Verbmobil's multi-blackboard and multi-engine architecture the results of concurrent processing threads can be combined in an incremental fashion. We argue that all results of concurrent processing modules must come with a confidence value, so that statistically trained selection modules can choose the most promising result. Packed representations together with formalisms for underspecification capture the uncertainties in each processing phase, so that the uncertainties can be reduced by linguistic, discourse and domain constraints as soon as they become applicable. Distinguishing features like the multilingual prosody module and the generation of dialog summaries are highlighted. We conclude that Verbmobil has successfully met the project goals with more than 80% of approximately correct translations and a 90% success rate for dialog tasks. One of the main lessons learned from the Verbmobil project is that the problem of speech-to-speech translation can only be cracked by the combined muscle of deep and shallow processing approaches.

1 Introduction

Verbmobil is a software system that provides mobile phone users with simultaneous dialog interpretation services for restricted topics [Wahlster, 1993; 2000b]. As the name Verbmobil suggests, the system supports **verb**al communication with foreign interlocutors in **mobil**e situations. It recognizes spoken input, analyses and translates it, and finally utters the translation. The multilingual system handles dialogs in three business-oriented domains, with bidirectional translation between three languages (German, English, and Japanese). In contrast to previous dialog translation systems that translate sentence-by-sentence, Verbmobil provides context-sensitive translations. Verbmobil uses an explicit dialog memory and exploits domain knowledge. The dialog context is used to resolve ambiguities and to produce an adequate translation in a particular conversational situation.

Figure 1: Mobile speech-to-speech translation with Verbmobil

Figure 1 illustrates the use of Verbmobil in a travel scenario. Let's suppose that an American business traveller has arrived at Frankfurt airport and wants to call Mrs. Meyer, the secretary of his German business partner. Since he does not speak German and knows that the secretary does not speak English, he activates Verbmobil using the voice dialing mode of his cell phone. After telling Verbmobil the phone number of Mrs. Meyer, the speech translation system initiates a conference call between the American traveller, the German secretary and Verbmobil. Verbmobil translates all input of the American speaker into German and all input of the German speaker into English.

Verbmobil is the first speech-only dialog translation system. Verbmobil users can simply pick up a standard mobile phone and use speech commands in order to initiate a dialog translation session (see Figure 2). The operation of the final Verbmobil system is completely hands-free without any push-to-talk button. Since the Verbmobil speech translation server can be accessed by GSM mobile telephones, the system can be used anywhere and anytime. No PC, notebook or PDA must be available to access the Verbmobil translation service, just a phone for each dialog participant. In addition, no waiting time for booting

computers and keyboard or mouse input to start the Verbmobil system is needed—dialog translation can begin instantaneously.

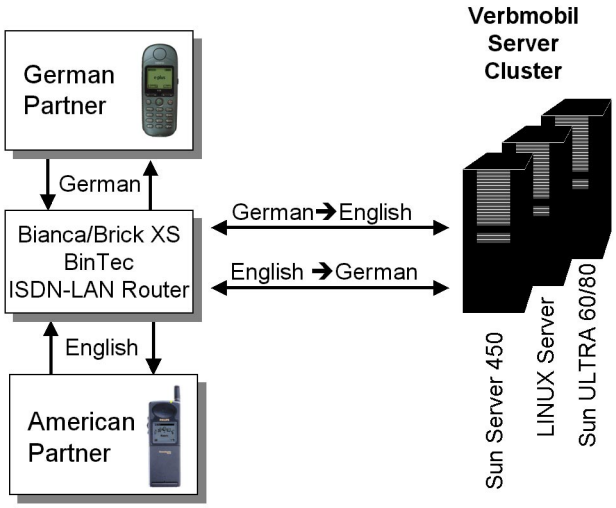

Figure 2: Three-party conference calls with Verbmobil

Verbmobil is the only dialog translation system to date based on an open microphone condition. It is not a "push-to-talk" system which has to be told which chunks of the sound signal represent coherent contributions by individual speakers: Verbmobil works that out for itself from the raw input signal. The signal may be of different qualities—not necessarily from a lab-quality close-speaking microphone, for instance it can be GSM (cell phone) quality. Thus, Verbmobil includes different speech recognizers for 16 kHz and 8 kHz sampling rates. Verbmobil is a speaker-adaptive system, i.e. for a new speaker it starts in a speaker-independent mode and after a few words have been uttered it improves the recognition results by adaptation. A cascade of unsupervised methods, ranging from very fast adaptation during the processing of a single utterance to complex adaptation methods that analyze a longer sequence of dialog turns, is used to adjust to the acoustic characteristics of the speaker's voice, the speaking rate, and pronounciation variants due to the dialectal diversity of the user community.

2 Understanding Spontaneous Speech

Verbmobil deals with spontaneous speech. This does not just mean continuous speech like in current dictation systems, but speech which includes realistic disfluencies and repair phenomena, such as changes of tack in mid-sentence (or mid-word), *ums* and *ers*, and cases where short words are accidentally left out in rapid speech. For example, in the Verbmobil corpus about 20% of all dialog turns contain at least one self-correction and 3% include false starts. Verbmobil uses a combination of shallow and deep analysis methods to recognize a speaker's slips and translate what he tried to say rather than what he actually said.

At an early processing stage prosodic cues are used to detect self-corrections. A stochastic model is used to segment the repair into the "wrong" part (the so-called reparandum) and the correction. Then the corrected input is inserted as a new hypothesis into the word hypotheses graph. Thus, Verbmobil's repair processing is a filter between speech recognition and syntactic analysis [Spilker et al, 2000]. The word lattice is augmented by an additional path that does no

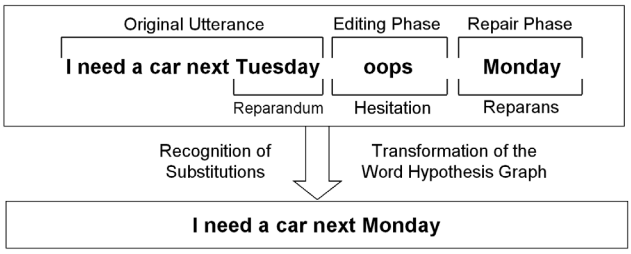

Figure 3: Repairing self-corrections

longer contain those parts of the utterances that the speaker tried to correct. This transformation of the word lattice is used in addition to simple disfluency filtering, that eliminates sounds like *ahh* that users often make while speaking (see Figure 3).

In addition to this shallow statistical approach, other forms of self-corrections are also processed at a later stage on the semantic level. A rule-based repair approach is applied during robust semantic processing to a chart containing possible semantic interpretations of the input (the so-called VIT Hypotheses Graph (VHG)). Verbmobil applies various hand-crafted rules to detect repairs in semantic representations and to delete parts of the representation that corresponds to slips of the speaker [Pinkal et al., 2000]

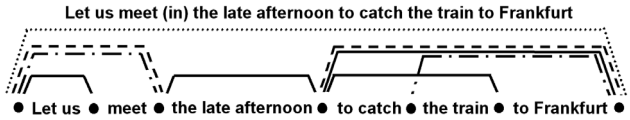

Figure 4: Finding a spanning analysis by type-raising

Due to a speech recognition error or a corrupted input signal, the word hypotheses graph in Figure 4 does not contain any temporal preposition in front of the temporal nominal phrase *the late afternoon*. A type coercion rule maps this phrase to a temporal modifier that expresses an underspecified temporal relation, that is later lexicalized as the default *in* during language generation.

Verbmobil deals with mixed-initiative dialogs between human participants. Each partner has a clear interaction goal

in a negotiation task like appointment scheduling or travel planning. Although these tasks encourage cooperative interaction, the participants have often conflicting goals and preferences that lead to argumentative dialogs. Therefore Verbmobil has to deal with a much richer set of dialog acts than previous systems that focused on information-seeking dialogs.

In order to ensure domain independence and scalability, Verbmobil was developed for three domains of discourse (appointment scheduling, travel planning, remote PC maintenance) with increasing size of vocabularies and ontologies. The travel planning scenario with a vocabulary of 10,000 words was used for the end-to-end evaluation of the final Verbmobil system. The PC maintenance task had a much larger vocabulary of almost 35,000 words from IT sublanguage lexica. Verbmobil is a hybrid system incorporating both deep and shallow processing schemes [Bub et al., 1997]. It integrates a broad spectrum of corpus-based and rule-based methods. Verbmobil combines the results of machine learning from large corpora with linguists' handcrafted knowledge sources to achieve an adequate level of robustness and accuracy.

3 Verbmobil's Training Corpora

A significant programme of data collection was performed during the Verbmobil project to extract statistical properties from large corpora of spontaneous speech. A distinguishing feature of the Verbmobil speech corpus is the multi-channel recording. The voice of each speaker was recorded in parallel using a close-speaking microphone, a room microphone, and various telephones (GSM phone, wireless DECT phone and regular phone), so that the speech recognizers could be trained on data sets with various audio signal qualities. The so-called partitur (German word for musical score) format used for the Verbmobil speech corpora orchestrates fifteen strata of annotations (see Figure 5, [Burger et al., 2000]). Multi-channel recordings of 3,200 spontaneous dialogs with 79,562 turns from 1,658 different speakers were transcribed and distributed on 56 CDs with a total of 21,5 GB of annotated speech corpora (available from BAS, see www.phonetik.uni-muenchen.de/Bas/BasKorporaeng.html).

In addition to the monolingual data, the multilingual Verbmobil corpus includes bilingual dialogs (from Wizard-of-OZ experiments, face-to-face dialogs with human interpreters, or dialogs interpreted by various versions of Verbmobil) and aligned bilingual transliterations. Three treebanks for German, English and Japanese have been developed with 85,000 trees annotated on three strata: morpho-syntax, phrase structure, and predicate-argument structure. The treebanks were used to train the statistical parser and the chunk parser. In addition, machine learning methods were applied to the treebanks to extract semantic construction rules and transfer rules for translation. The end-to-end evaluations of the various Verbmobil prototypes have shown clearly, that the robustness, coverage, and accuracy of a speech-to-speech translation system for spontaneous dialogs depends critically on the quantity and quality of the training corpora.

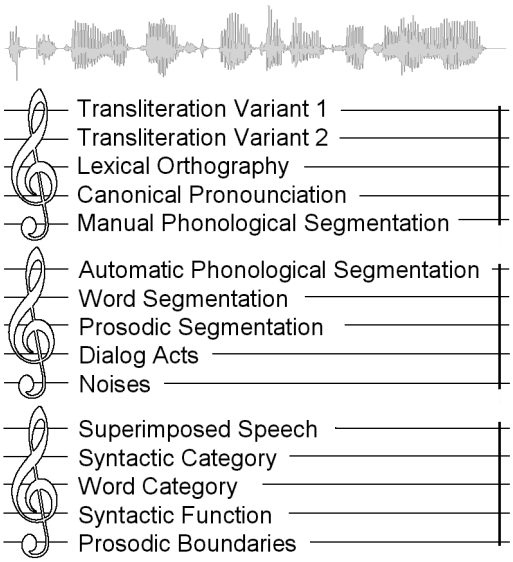

Figure 5: Verbmobil's multi-stratal annotation of speech

4 The Anatomy of Verbmobil

A distinguishing feature of Verbmobil is its multi-engine parsing and translation architecture. The screenshot of Verbmobil's control panel provides an overview of the main components of the system (see Figure 6). The overall control and data flow is indicated by arrows pointing upwards on the left side of the screenshot, from left to right in the middle and downwards on the right side. On the bottom various input devices can be selected. Since Verbmobil is a multilingual system it incorporates four speech recognizers and four speech synthesizers for German, English, Japanese, and Chinese.

Three parsers based on different syntactic knowledge sources are used to process the word hypotheses graphs (WHG) that are augmented by prosodic information extracted by the prosody module (see Section 5 below). All parsers use the multi-stratal VIT representation as an output format. VITs (Verbmobil Interface Terms) are used as a multi-stratal semantic representation by the central blackboards for the deep processing threads in Verbmobil. The semantic representation in a VIT is augmented by

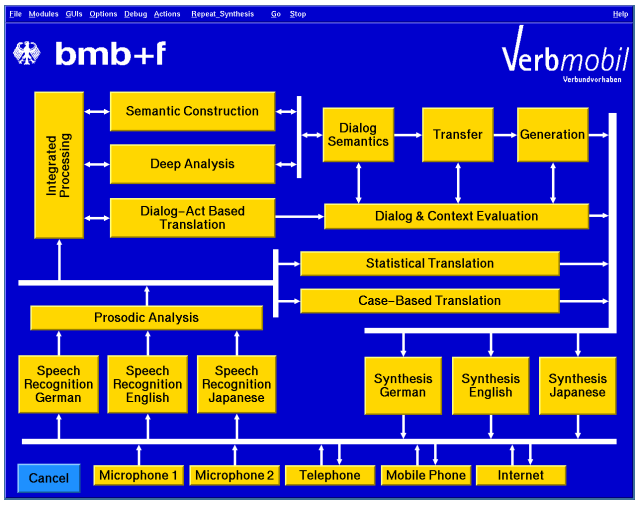

Figure 6: A snapshot of Verbmobil's control panel

various features concerning morpho-syntax, tense, aspect, prosody, sortal restrictions and discourse information. VITs form the input and output of the modules for robust semantic processing and semantic-based transfer. The initial design of the VIT representation language was inspired by underspecified discourse representation structures (UDRS, [Reyle, 1993]. VITs provide a compact representation of lexical and structural ambiguities and scope underspecification of quantifiers, negations and adverbs. The linguistic information is encoded into variable-free sets of non-recursive terms (see Figure 7). These streams of literals serve as flat multi-stratal representations that are very efficient for incremental processing. The various linguistic strata are cross related by a labelling system. Since VIT terms are the central information structure in Verbmobil, they are treated as an abstract data type. VITs are used as a common representation scheme for linguistic information exchange between all components and processing threads of Verbmobil.

Since in most cases the parsers produce only fragmentary analyses, their results are combined in a chart of VIT structures. A chart parser and a statistical LR parser are combined in a package that is visualized in the screenshot as "integrated processing". These shallow parsers produce trees that are transformed into VIT structures by a module called semantic construction (see Figure 6). This syntax-semantics interface is primarily lexically driven [Schiehlen 2000]. The module with the label "deep analysis" is based on a HPSG parser for deep linguistic processing in the Verbmobil system. Verbmobil is the only completely operational speech-to-speech translation system that is based on a wide-coverage unification grammar and tries to preserve the theoretical clarity and elegance of linguistic analyses in a very efficient implementation. The parser for the HPSG grammars processes the n best paths produced by the integrated processing module. It is implemented as a bidirectional bottom-up active chart parser [Kiefer et al., 2000]

```
Vit (vitID (sid (104,a,en,10,80,1,en,y,semantics), % SegmentID
          [word (he, 1, [26]),                     % WHG String
           word(is, 2, []),
           word(coming, 3, [27]),
           word(at, 4, [36]),
           word(the ,5, [28]),
           word(beginning, 6, [35]),
           word(of, 7, [35]),
           word(``August'', 8, [34])]),
     index (38, 25 ,i35),                          % Index
    [beginning (35, i37),                          % Conditions
     arg3 (35, i37 ,i38),
     come (27, i35),
     arg1 (27, i35, i36),
     decl (37, h43),
     pron (26, i36),
     at (36, i35, i37),
     mofy (34 ,i38, aug),
     def (28, i37, h42, h41),
     udef (31, i38, h45, h44)],
    [in_g (26, 25), in_g (37, 38),                 % Constraints
     in_g (27, 25), in_g (28, 30),
     in_g (31, 33), in_g (34, 32),
     in_g (35, 29), in_g (36, 25),
     leq (25, h41), leq (25, h43),
     leq (29, h42), leq (29, h44),
     leq (30, h43), leq (32, h45),
     leq (33, h43)],
    [s_sort (i35, situation),                      % Sorts
     s_sort (i37, time),
     s_sort (i38, time)],
    [dialog_act (25, inform),                      % Discourse
     dir (36, no),
     prontype (i36, third,std)],
    [cas (i36, nom),                               % Syntax
     gend (i36, masc),
     num (i36, sg), num (i37, sg), num (i38, sg),
     pcase (l135, i38, of)],
    [ta_aspect (i35, progr),                       % Tense and Aspect
     ta_mood (i35, ind),
     ta_perf (i35, nonperf),
     ta_tense (i35, pres)],
    [pros_accent (l135)]                           % Prosody
```

Figure 7: VIT for "He is coming at the beginning of August"

The statistical translation module starts with the single best sentence hypothesis of the speech recognizer [Vogel et al., 2000]. Prosodic information about phrase boundaries and sentence mode are utilized by the statistical translation module. The output of this module is a sequence of words in the target language together with a confidence measure that is used by the selection module (not shown in the control panel) for the final choice of a translation result. Verbmobil includes two components for case-based translation. Substring-based translation is a method for incremental synchronous

interpretation, that is based on machine learning methods applied to a sentence-aligned bilingual corpus. Substrings of the input for which a contiguous piece of translation can be found in the corpus are the basic processing units. Substring pairs are combined with patterns for word order switching and word cluster information in an incremental translation algorithm for a sequence of input segments [Block, 2000]. The other component for case-based translation is based on 30,000 translation templates learned from a sentence-aligned corpus. Date, time and naming expressions are recognized by definite clause grammars (DCGs) and marked in the WHG. An A* search explores the cross-product graph of the WHG with the subphrase tags and the template graph. A DCG-based generator is used to produce target language output from the interlingual representation of the recognized date, time and naming expressions. These subphrases are used to instantiate the target language parts of translation templates.

Dialog-act based translation includes the statistical classification of 19 dialog acts and a cascade of more than 300 finite-state transducers that extract the main propositional content of an utterance. The statistical dialog classifier is based on n-grams and takes the previous dialog history into account. The recognized dialog act, the topic and propositional content are represented by a simplistic frame notation including 49 nested objects with 95 possible attributes covering the appointment scheduling and travel planning tasks. A template-based approach to generation is used to transform these interlingual terms into the corresponding target language. The shallow interlingual representation of an utterance is stored together with topic and focus information as well as a deep semantic representation encoded as a VIT in the dialog memory for further processing by the dialog and context evaluation component.

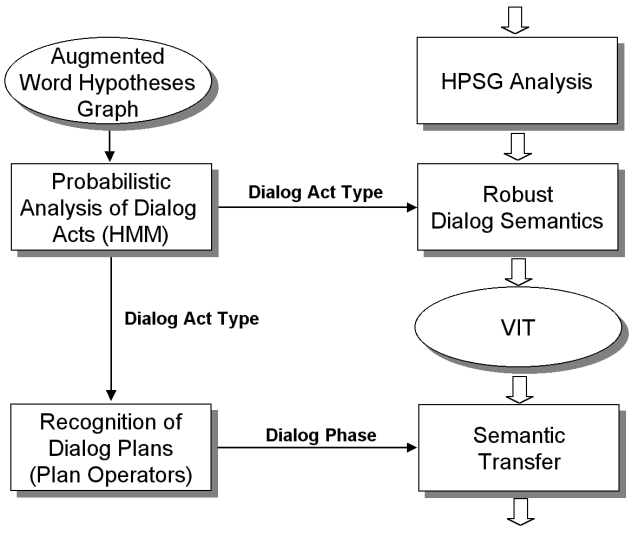

Figure 8: The use of stochastic dialog act and plan recognition

The dialog component includes a plan processor, that structures an ongoing dialog hierarchically in different dialog phases, games and moves. Dialog acts are the terminal nodes of the tree structure that represents the dialog structure. Information about the dialog phase is used e.g. during the semantic-based transfer for disambiguation tasks (see Figure 8). In addition, inference services are provided by the dialog and context component eg. for the completion of underspecified temporal expressions and the resolution of anaphora or ellipsis. Temporal reasoning is used for example to transform expressions like *two hours later* or *next week* into fully specified times and dates stored in the dialog memory for summarizing the results of a negotiation. The transfer module triggers contextual reasoning process only in cases where a disambiguation or resolution is necessary for a given translation task. For example, the German noun *Essen* can be translated into *lunch* or *dinner* depending on the time of day, which can be derived by contextual reasoning. Disambiguation and resolution on demand is typical for Verbmobil's approach to translation, since various forms of underspecification and ambiguity can be carried over into target language, so that the hearer can resolve them. Consider the German sentence *Wir treffen die Kollegen in Berlin* and its English equivalent *We will meet the colleagues in Berlin*. English and German have the same PP-attachment ambiguity in which *in Berlin* is either attached to the noun phrase *the colleagues* or to the verb *meet*.

The transfer component is basically a rewriting system for underspecified semantic representations using Verbmobil's VIT formalism [see Emele et al., 2000]. Semantic-based transfer receives a VIT of a source language utterance and transforms it into a VIT for the target language synthesis. This means that the transfer module abstracts away from morphological and syntactic analysis results. The final Verbmobil system includes more than 20,000 transfer rules. These rules include conditions that can trigger inferences in the dialog and context evaluation module to resolve ambiguities and deal with translation mismatches, whenever necessary. The transfer component uses cascaded rule systems, first for the phrasal transfer of idioms and other non-compositional expressions and then for the lexical transfer. The translation of spatial and temporal prepositions is based on an interlingual representation in order to cut down the number of specific transfer rules. Semantic-based transfer is extremely fast and consumes on the average less than 1% of the overall processing time for an utterance.

Verbmobil's multilingual generator includes a constraint-based microplanning component and a syntactic realization module that is based on the formalism of lexicalized tree-adjoining grammars [see Becker et al , 2000]. The input to the microplanning component are VITs produced by the transfer module. A sentence plan is generated that consists basically of lexical items and semantic roles linking them together. The microplanner decides about subordination, aggregation, focus

and theme control as well as anaphora generation. The syntactic realization component can either use LTAG grammars that are compiled from the HPSG grammars used for deep analysis or a hand-written LTAG generation grammar. For English and Japanese the grammars that were designed for analysis are usable for generation after an offline-compilation step.

The speech synthesizer for German and American English follows a concatenative approach based on a large corpus of annotated speech data. The word is the basic unit of concatenation, so that subword units are only used if a word is not available in the database.

The synthesizer applies a graph-based unit selection procedure to choose the best available synthesis segments matching the segmental and prosodic constraints of the input. Whenever possible the synthesizer exploits the syntactic, prosodic and discourse information provided by previous processing stages. Thus for the deep processing stream it provides concept-to-speech synthesis, whereas for the shallow translation threads it operates more like a traditional text-to-speech system resulting in a lower quality of its output.

Another novel functional feature of Verbmobil is the ability to generate dialog summaries. Suppose that two speakers negotiate a travel plan: one can ask the system either to specify the final agreement, omitting the negotiating steps, or to summarize the steps of argument while leaving out irrelevant details of wording. A dialog summary can be produced on demand after the end of a conversation.

The summaries are based on the semantic representation of all dialog turns stored in the dialog memory of Verbmobil. It is interesting to note that dialog summaries are mainly a bye-product of the deep processing thread and the dialog processor of Verbmobil. The most specific accepted negotiation results are selected from the dialog memory [Alexandersson et al., 2000]. The semantic-based transfer component and the natural language generators for German and English are used for the production of multilingual summaries. This means that after a conversation over a cell phone the participants can ask for a written summary of the dialog in their own language. The dialog summary can be sent as an HTML document using email. In the context of business negotiations Verbmobil's ability to produce written dialog summaries of a phone conversation is an important value-added service.

5 Exploiting Prosodic Information

Verbmobil is the first spoken-dialog interpretation system that uses prosodic information systematically at all processing stages. The results of Verbmobil's multilingual prosody module are used for parsing, dialog understanding, translation, generation and speech synthesis (see Figure 9). This means that prosodic information in the source utterance is passed even through the translation process to improve the generation and synthesis of the target utterance. Prosodic differences in one language can correspond to lexical or syntactic differences in another; for instance, a German utterance beginning *wir haben noch ...* may be translated by Verbmobil into English either as *we still have ...* or as *we have another ...* depending whether *noch* is stressed. Although prosody is used in some other recent speech recognition systems, the exploitation of prosodic information is extremely limited in these approaches. For example, the ATR Matrix system [see Takezawa et al., 1998] uses prosody only to identify sentence mood (declarative vs. question). We believe that Verbmobil is the first fully operational system to make significant use of prosodic aspects of speech.

The prosody module of Verbmobil uses the speech signal and the word hypotheses graph (WHG) produced by the speech recognizer as an input and outputs an annotated WHG with prosodic information for each recognized word. The system extracts duration, pitch, energy, and pause features and uses them to classify phrase and clause boundaries, accented words and sentence mood. A combination of a multilayer perceptron and a polygram-based statistical language model annotate the WHG with probabilities for the classified prosodic events.

Verbmobil uses the probabilistic prosodic information about clause boundaries to reduce the search space for syntactic analysis dramatically. During parsing, the clause boundary marks that are inserted into the WHG by the prosody module play the role of punctuation marks in written language. Dialog act segmentation and recognition is also based on the boundary information provided by the prosody module. Prosodic cues about sentence mood is often used in Verbmobil's translation modules to constrain transfer results, if there is not enough syntactic or semantic evidence for a certain mood (e.g. question). The information about word accent is used to guide lexical choice in the generation process. Finally, during speech synthesis the extracted prosodic features are used for speaker-adaptation.

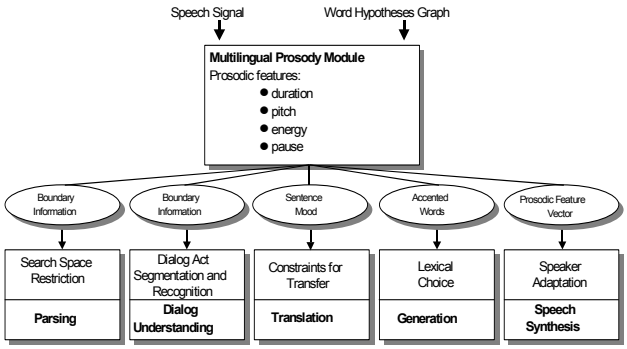

Figure 9: The role of prosodic information in Verbmobil

6 Verbmobil's Multi-Blackboard Architecture

The final Verbmobil system consists of 69 highly interactive modules. The transformation of speech input in a source language into speech output in a target language requires a tremendous amount of communication between all these modules. Since Verbmobil has to translate under real-time conditions it exploits parallel processing schemes whenever possible. The non-sequential nature of the Verbmobil architecture implies that not only inputs and results are exchanged between modules but also top-down expectations, constraints, backtracking signals, alternate hypotheses, additional parameters, probabilities, and confidence values.

198 blackboards are used for the necessary information exchange between modules. A module typically subscribes to various blackboards. Modules can have several instances, e.g. in a multiparty conversation there may be two German speakers, so that two instances of the German speech recognition module are needed.

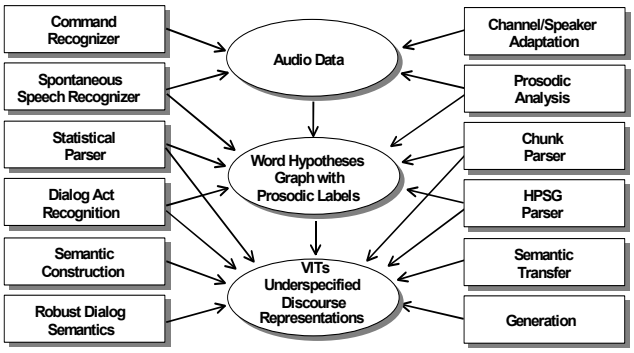

Figure 10**:** Some key blackboards with their subscribing modules

The final Verbmobil system is based on a multi-blackboard architecture that pools processing modules around blackboards representing intermediate results at each processing stage (see Figure 10). It turned out that such a multi-blackboard approach is much more efficient than the more general multi-agent architecture used in the first Verbmobil prototype. Due to the huge amount of interaction between modules a multi-agent architecture with direct communication among module agents would imply 2380 different interfaces for message exchanges between the 69 agents.

In a multi-blackboard architecture based on packed representations at all processing stages (speech recognition, parsing, semantic processing, translation, generation, speech synthesis) using charts with underspecified representations the results of concurrent processing threads can be combined in an incremental fashion. All results of concurrent processing modules come with a confidence value, so that selection modules can choose the most promising results at each processing stage or delay the decision until more information becomes available. Packed representations such as the WHG (Word Hypotheses Graph) and VHG (VIT Hypotheses Graph) together with formalisms for underspecification capture the non-determinism in each processing phase, so that the remaining uncertainties can be reduced by linguistic, discourse and domain constraints as soon as they become applicable.

7 Verbmobil's Multi-Engine Approach

Verbmobil performs language identification, parsing and translation with several engines simultaneously. Whereas the multi-engine parsing results are combined and merged into a single chart, a statistical selection module chooses between the alternate results of the concurrent translation threads, so that only a single translation is used for generating the system's output.

Verbmobil uses three parallel parsing threads: an incremental chunk parser, a probabilistic LR parser and a HPSG parser. These parsers cover a broad spectrum with regard to their robustness and accuracy. The chunk parser [Hinrichs et al., 2000] produces the most robust but least accurate results, whereas the HPSG parser delivers the most accurate but least robust analysis. All parsers process the same word hypotheses graph with its prosodic annotations. The search for the best scored path (according to the acoustic score and the language model) is controlled by a central A* algorithm that guides the three parsers through the word hypotheses graph. The HPSG parser may return more than one analysis for ambiguous inputs, whereas the chunk parser and statistical parser return always only one result. Each parser uses a semantic construction component to transform its analysis results into a semantic representation term.

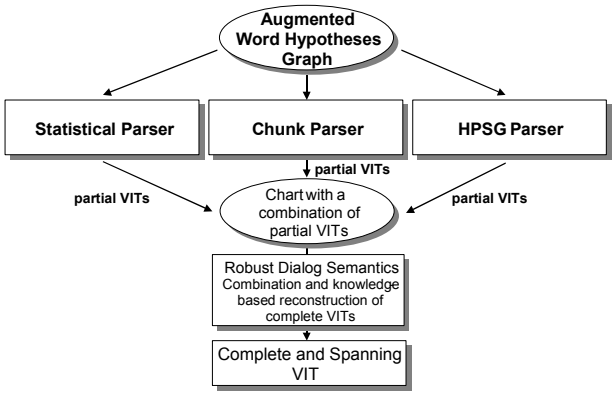

Figure 11: Verbmobil's multi-engine parsing approach

Even partial results of the different parsing engines are integrated into a chart of VITs, that is further analyzed by the robust semantic processing component (see Figure 11).

The final Verbmobil system includes five translation engines (see Figure 12): statistical translation, case-based translation, substring-based translation, dialog-act based translation, and semantic transfer. These engines cover a wide spectrum of translation methods. While statistical translation is very robust against speech recognition problems and produces quick-and-dirty results, semantic transfer is computationally more expensive and less robust but produces higher quality translations. However, it is one of the fundamental insights gained from the Verbmobil project, that the problem of robust, efficient and reliable speech-to-speech translation can only be cracked by the combined muscle of deep and shallow processing approaches.

The translation quality of the final Verbmobil system was rigorously evaluated. 65 evaluators checked 43,180 Verbmobil translations and judged their correctness. We call a translation "approximately correct", if it preserves the intention of the speaker and the main information of his utterance. Table 1 shows clearly that no single translation engine achieves more than 81% approximately correct translations, but that the selection of the appropriate translation result increases the overall performance significantly. In Verbmobil, we used the judgements of the human evaluators (see Table 1, Manual Selection) to construct a training corpus for an instance-based learning algorithm that picks the best translation for a given WHG of a particular turn segment (see Table 1, Automatic Selection).

Translation Thread	Word Accuracy ≥ 50% 5069 Turns	Word Accuracy ≥ 75% 3267 Turns	Word Accuracy ≥ 80% 2723 Turns
Case-based Translation	37%	44%	46%
Statistical Translation	69%	79%	81%
Dialog-Act based Translation	40%	45%	46%
Semantic Transfer	40%	47%	49%
Substring-based Translation	65%	75%	79%
Automatic Selection	**78%**	**83%**	**85%**
Manual Selection	**88%**	**95%**	**97%**

Table 1: Quality of Translations from German to English

The language identification component of Verbmobil uses also a multi-engine approach to identify each user's input language. The three instances of the multilingual speech recognizer for German, English, and Japanese run concurrently for the three first seconds of speech input. A confidence measure is used to decide which language is spoken by a particular dialog participant. The language identification component switches to the selected recognizer that produces a word hypotheses graph for the full utterance. Verbmobil's error rate for this type of language identification task is only 7.3% [see Waibel et al., 2000]. Verbmobil's architecture supports multiple process instances of all components, so that Verbmobil can be used as a translation server for multiparty dialogs (e.g. two Germans, a Japanese and an American planning a joint trip).

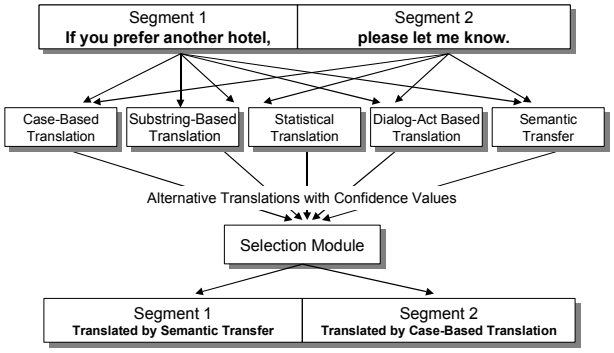

Figure 12. The multi-engine translation approach of Verbmobil

8 Lessons Learned from Verbmobil

The broad range of scientific discoveries in the areas of speech, language and discourse processing, dialog translation, language generation and speech synthesis that resulted from the Verbmobil project are documented in more than 800 publications (www.verbmobil.de) and a comprehensive book [Wahlster, 2000a].

One of the main lessons learned from the Verbmobil project is that the problem of speech-to-speech translation of spontaneous dialogs can only be cracked by the combined muscle of deep and shallow processing approaches:

- deep processing can be used for merging, completing and repairing the results of shallow processing strategies
- shallow methods can be used to guide the search in deep processing
- statistical methods must be augmented by symbolic models to achieve higher accuracy and broader coverage
- statistical methods can be used to learn operators or selection strategies for symbolic processes

The final Verbmobil architecture supports large and robust dialog systems and maximizes the necessary interaction between processing modules:

- in Verbmobil's multi-blackboard and multi-engine architecture, that is based on packed representations on all processing levels and uses charts with underspecified multi-stratal representations, the results of concurrent processing threads can be combined in an incremental fashion
- all results of concurrent and competing processing modules come with a confidence value, so that statistically trained selection modules can choose the most promising result at each stage, if demanded by a following processing step.

– packed representations together with formalisms for underspecification capture the uncertainties in each processing phase, so that the uncertainties can be reduced by linguistic, discourse and domain constraints as soon as they become applicable. In particular, underspecification allows disambiguation requirements to be delayed until later processing stages where better-informed decisions can be made.
– The massive use of underspecification makes the syntax-semantic interface and transfer rules almost deterministic, thereby boosting processing speed.

Verbmobil has shown the need to take software engineering considerations in language technology projects seriously. Verbmobil's system integration group included professional software engineers with no particular language or speech technology background; they were responsible for ensuring that the software is robust and maintainable, and that modules developed in different programming languages by a distributed team fit together properly. These issues are too important to leave to subject specialists who see them as a side issue. An important achievement of the Verbmobil project is the consistent integration of a very large number of modules created by diverse groups of researchers from disparate disciplines and to produce a set of capabilities which have not been demonstrated in an integrated speech-to-speech translation system before.

Organizationally, Verbmobil underlines the importance of competition among research teams, with frequent objective evaluations. Competition was fostered naturally within the Verbmobil framework, because the processing model itself is a competitive one. Crucial to the success of Verbmobil was the fact that various teams within the project developed rival solutions to particular tasks, with formal evaluations being used to winnow out the most successful or to combine it with the next best solutions to improve the overall performance of the system.

The objective of the public funding provided by the German Federal Ministry of Education and Research (BMBF) for the Verbmobil Project has been to bring European language technology to the stage of achieving real industrial impact by the turn of the century. Participating companies developing spin-off applications at their own expense have already brought twenty products to market that are all based on results from Verbmobil. There have been various patents and inventions resulting from Verbmobil, in areas such as speech processing, parsing, dialog, machine translation and generation. Seven spin-off companies in language technology have been created by former Verbmobil researchers. For example, AixPlain (www.aixplain.de) markets speech translation systems, Sympalog (www.sympalog.de) develops spoken dialog systems, and XtraMind (www.xtramind.com) delivers email response systems based on Verbmobil technology. At present, Verbmobil's large industrial partners (DaimlerChrysler, Philips, Siemens, Temic) are among the top European companies using language technology in the marketplace.

The sharable language resources collected and distributed during the Verbmobil project will be useful beyond the project lifespan, since these transliterated and richly annotated corpora of spontaneously spoken dialogs can be used for building, improving or evaluating natural language and speech algorithms or systems in coming years.

Along the way, the Verbmobil project has done a great deal to bring researchers in Germany together across the language/speech and the academic/industrial divides. This is an important contribution from the point of view of a long-range research policy for the field of human language technology. More than 900 young researchers (among them 238 master students, 164 PhD students, and 16 habilitation postdocs) gained experience in advanced speech and language technology through their work on Verbmobil during the project lifespan.

9 Conclusion and Future Work

Although Verbmobil was a high-risk and long-term project (1993 – 2000), it has successfully met its technical project goals. The Verbmobil consortium brought together 31 partners across three continents. The total amount of public and private funding was about $80 million, resulting in Europe's largest AI project. The technical challenge was to design and implement

- a speaker-independent and bidirectional speech-to-speech translation system for spontaneous dialogs in mobile situations
- that works in an open microphone mode and can cope with speech over GSM mobile phones
- for four language pairs, three domains and a vocabulary size of more than 10,000 word forms
- with an average processing time of four times of the input signal duration
- with a word recognition rate of more than 75% for spontaneous speech
- with more than 80% of approximately correct translations that preserve the speaker's intended effect on the recipient in a large-scale translation experiment
- a 90% success rate for dialog tasks in end-to-end evaluations with real users

Various benchmark tests and large-scale end-to-end evaluation experiments with unseen test data have convincingly shown that all these objectives have been met by the final Verbmobil system and some goals have been surpassed [Tessiore and v. Hahn, 2000].

SmartKom (1999-2003) is the follow-up project to Verbmobil and reuses some of Verbmobil's components for

the understanding of spontaneous dialogs. SmartKom is a multimodal dialog system that combines speech, gesture, and mimics input and output [Wahlster et al, 2001]. Spontaneous speech understanding is combined with the video-based recognition of natural gestures. One of the major scientific goals of SmartKom is to design new computational methods for the seamless integration and mutual disambiguation of multimodal input and output on a semantic and pragmatic level. SmartKom is based on the situated delegation-oriented dialog paradigm (SDDP), in which the user delegates a task to a virtual communication assistant, visualized as a life-like character on a graphical display. The main contractor of the SmartKom consortium is the German Research Center for Artificial Intelligence (DFKI). The major industrial partners involved in SmartKom are DaimlerChrysler, Philips, Siemens and Sony.

References

[Alexandersson et al., 2000], Jan Aleandersson., Peter Poller, and Michael Kipp, Generating Multilingual Dialog Summaries and Minutes. In: [Wahlster, 2000a] 507-518.

[Becker et al, 2000] Tilman Becker, Anne Kilger, Patrice Lopez, and Peter Poller. The Verbmobil Generation Component VM-GECO. In [Wahlster, 2000a], 481-496.

[Block, 2000] Hans Ulrich Block. Example-Based Incremental Synchronous Interpretation. In [Wahlster, 2000a], 411-417.

[Bub et al., 1997] Thomas Bub, Wolfgang Wahlster, and Alex Waibel, A. Verbmobil: The Combination of Deep and Shallow Processing for Spontaneous Speech Translation. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing, München, Germany*, 71–74, IEEE *1997*.

[Burger et al, 2000] Susanne Burger, Karl Weilhammer, Florian Schiel, and Hans G. Tillmann. Verbmobil Data Collection and Annotation. In [Wahlster, 2000a], 537-549.

[Emele et al., 2000] Martin Emele, Micheal Dorna, Anke Lüdeling, Heike Zinsmeister, and Christian Rohrer. Semantic-Based Transfer. In [Wahlster, 2000a], 359-376.

[Hinrichs et al., 2000] Erhard Hinrichs, Sandra Kübler, Valia Kordoni, and Frank Müller. Robust Chunk Parsing for Spontaneous Speech. In [Wahlster, 2000a], 163-182.

[Kiefer et al., 2000] Bernd Kiefer, Hans-Ulrich Krieger, and Mark Jan Nederhof. Efficient and Robust Parsing of Word Hypotheses Graphs. In [Wahlster, 2000a], 279-295.

[Pinkal et al., 2000] Manfred Pinkal, C.J. Rupp, and Karsten Worm. Robust Semantic Processing of Spoken Language. In [Wahlster, 2000a], 321-335.

[Reyle, 1993] Uwe Reyle. Dealing with Ambiguities by Under-specification: Construction, Representation and Deduction. Journal of Semantics 10 (2): 123-179, 1993.

[Schiehlen, 2000] Michael Schiehlen. Semantic Construction. In [Wahlster, 2000a], 200-215.

[Spilker et al., 2000] Jörg Spilker, Martin Klarner, and Günther Görz. Processing Self-Corrections in a Speech-to-Speech System. In [Wahlster, 2000a], 131-140.

[Takezawa et al., 2000] Toshiyuki Takezawa, Tsuyoshi Morimoto, Yonishori Sagisaka, Nick Campbell, Hitoshi Iida, Fumiaki Sugaya, Akio Yokoo, and Seiichi Yamamoto. A Japanese-to-English Speech Translation System: ATR-MATRIX. In *Proceedings of the 5th International Conference on Spoken Language Processing, ICSLP*, Sydney, 957–960, 1998.

[Tessiore and v. Hahn, 2000] Lorenzo Tessiore and Walther v. Hahn. Functional Validation of a Machine Interpretation System: Verbmobil. In [Wahlster, 2000a], 611-631.

[Vogel et al, 2000] Stepham Vogel, Franz Josef Och, Christoph Tillmann, Sonja Niessen, Hassan Sawaf, and Hermann Ney. Statistical Methods for Machine Translation. In [Wahlster, 2000a], 377-393.

[Waibel et al, 2000] Alex Waibel, Hagen Soltau, Tanja Schultz, Thomas Schaaf, and Florian Metze. Multilingual Speech Recognition. In [Wahlster, 2000a], 33-45.

[Wahlster, 1993] Wolfgang Wahlster. Verbmobil: Translation of Face-to-Face Dialogs. In *Proceedings of the Fourth Machine Translation Summit*, Kobe, Japan, 128–135.

[Wahlster, 2000a] Wolfgang Wahlster (ed.). Verbmobil: Foundations of Speech-to-Speech Translation. Berlin, New York: Springer 2000.

[Wahlster, 2000b] Wolfgang Wahlster. Mobile Speech-to-Speech Translation: An Overview of the Final Verbmobil System. In: [Wahlster, 2000a], 3-21.

[Wahlster et al., 2001] Wolfgang Wahlster, Norbert Reithinger, and Anselm Blocher. SmartKom: Multimodal Communication with a Life-Like Character. In: *Proceedings of the 7th European Conference on Speech Communication and Technology*, Aalborg, Eurospeech, 2001.