GETTING GOOD VALUE

Facts, Values, and Goals In Computational Linguistics *

Michael A. Gilbert

Department of Philosophy, York University, Toronto, Canada
Gilbert@YorkU.ca

Abstract. This discussion is intended to amplify the role of various intentional components in computer-human communication. In attempting to further the degree to which human subjects and computers can disagree and communicate within the context of that disagreement, i.e., "argue," the ability to identify and classify various locutions as facts, values, or goals is crucial. I suggest that inquiry into the conceptual framework, context, the identification of pre-existing beliefs, values and goals, and the discussion of their priority will facilitate the sought after communication. Toward this end, I use the concept of 'field' and its identification to help distinguish the relevant categories.

The distinction between facts and values has bedeviled philosophers through the ages. Plato pointed out that facts are the things we never argue about, we just go and find out what they are. Aristotle distinguished between *logos, pathos* and *ethos* with the first having far more rigorous aspects than the latter two. His "practical" syllogisms have been every bit as fraught with contention and uncertainty as his purely deductive have been precise and unquestioned. And it was David Hume who emphasized that one cannot move from any number of facts to even the simplest value. What became forever known as the "Is-Ought Problem" stated that "is" statements, i.e., statements of fact, cannot lead, deductively, to statements of value unless a normative axiom or postulate is put in. Thus, for example, the myriad of facts about how kind grandmother has been to my son cannot, *in and of themselves*, lead to the conclusion that he ought telephone her more often. One inevitably needs the moral principle that, say, you ought be nice to those who are nice to you, to make the normative leap.

In the last century the most celebrated usage of the fact-value distinction was by the logical positivists. Logical Positivism, coming into its own as it did immediately after World War I, was in no small part a reaction to the metaphysical idealism, religious emphasis, and unreasoning nationalism of the time. The Logical Positivists felt that the saving of humanity required the abandonment of mysticism, religion, and sentiment as guiding principles and the adoption, in their place, of scientific principles and reasoning. [1]. The heart of their programme

* This work has been supported in part by Social Sciences and Humanities Research Council Grant Nr. 5 512790. I also want to express my debt to the referees who were generally very helpful to and tolerant of an outsider.

was the adoption of the Principle of Verifiability [PV] which, in its essence was exquisitely simple: Those statements that had, *in principle*, a means of verification were factual and had truth values, while those that did not were "mere" poetry or expressions of emotion. As a result,

1. [$6 + 7 = 13$]

is a fact, while

2. "13 is a lucky number,"

comes down to

3. "I like 13."

Similarly,

4. "The CN tower in Toronto is 553.33m meters high,"

or

5. "The CN Tower is the world's tallest free standing structure,"

are facts and, therefore, true or false, while the statements,

6. "The CN Tower is an engineering marvel,"

or

7. "The CN Tower is a dreadful waste of money,"

are expressions of feeling and neither true nor false.

Logical Positivism met a very unusual fate for a philosophical theory: it simply died. The death blow came from Gödel's theorem, but other problems were already weakening the foundations. Nonetheless, the underlying premiss that there is an identifiable distinction that can be made between facts and values has had a major impact.

The real difficulty, however, goes beyond the issue of the ascendancy of science over poetry to the very nature of actual decidability itself. Certainly, (1) is true, but someone might object that, in fact, [$6 + 7$] really equals D because the hexadecimal system is essentially more basic than the metric. Also, we may agree that nothing is more basic than [$1 + 1 = 2$], but adding one drop of water to another drop of water yields one drop of water. Clearly, this is a sophism, but a telling one nonetheless.

Consider (4). As it is a classically empirical statement it is either true or false. We know how to determine the height of the Tower, so the statement expressing its height is a fact. But anyone really being careful about the statement will readily admit that a variety of factors will enter into the determination of the truth and falsity of (4). Factors such as the tools or methods used for measurement, the outside temperature, season and time of day might all result in different measurements. Can something as high and complex as the CN Tower be measured once and for all, and, if so, then to what degree of accuracy? Well, for the guidebooks, it is not a problem, but for other purposes the assumption that the measurement can be accurate to .33 meters without a *ceteris paribus* is far fetched.

An alternate attack comes from precisely the opposite direction. While (7) may well be highly contentious and a question of values, what about (6)? Surely, one wants to say, "if anything is an engineering marvel, then the CN Tower is." Is the fact that we do not have scientific criteria for what constitutes an engineering marvel sufficient grounds for saying that (6) only has meaning as an expression of emotion? Not only that, but such statements have logical consequences and corollaries as well as identifiable presumptions that can be laid out just as clearly as measurement techniques.

The difficulty that brought down the PV was that the very distinctions on which it rested – empirical vs. non-empirical, testable vs. non-testable – were fraught with hidden theoretic assumptions without which tests and observations could not even begin. The philosophical approaches that succeeded Logical Positivism in the 1940s and '50s were language based and emphasized the role of language in both epistemology and metaphysics. Subsequently, in the 70s and 80s the impact of the philosophy of science, especially as characterized by Popper, Kuhn, Feyerabend [10, 8, 2] and others began to have an enormous influence on philosophical work. These views disdained, to one degree or another, the idea that facts are independently identifiable, independent, that is, of the theory in which they are identified. Rather, it was argued, the theory itself determines what is and is not acceptable (even existent) data, as well as determining what are the correct and legitimate modes of testing, confirmation, and measurement.

Obviously, if the identification of facts and values are dependent upon the theory being used, then a machine's ability to distinguish between them will run into difficulties. The Logical Positivist approach of identifying by means of testing falls apart if the very system being examined is what determines the means of testing. From the point of view of computational linguistics, it is not so much the issues raised in the philosophy of science that are telling, as the difficulties inherent in interactions where the category of statement is not well identified. But this, if embraced, may very well be the solution as opposed to the problem.

To say that the distinction between facts and values is theory dependent, and to accept that what is true, what is false, and how these are tested and separated is also determined by the very theory that might be under discussion might well point to a solution of sorts. Namely, the acceptance that there is no real theory independent difference between facts and values. That is, all truths, all facts and/or values, are relative to a theory, and, most importantly, *every user is a theory.*

If we think in terms of science and mathematics, then a theory is something quite precise and delimited. There are axioms, postulates, corollaries, consequences and a plethora of background assumptions. But a machine interacting with an individual is working with something much less precise. Certainly, an individual person may hold any number of theories (not all of which may be consistent one with another,) but there will also be a multitude of non-theoretic and even non-reflective beliefs, values and goals that come along with the person. When dealing with persons the best approach is not to think in terms of theories, but of fields. Fields range from those that one inherits such as race, religion and native culture, to those that are learned and acquired such as occupation, individual interests and avocations. Every field brings with it assumptions regarding the world and what is valued in it. Very importantly, fields bring with them goals, priorities and values. Identifying an individual's field memberships is one way of identifying her/his goals and values. The term was first introduced by Toulmin [11] to account for those aspects of reasoning that varied from domain to domain, and those that did not. Eligible evidence, for example, varied according to fields as when, to cite an extreme example, biblical scripture citation is acceptable in one sort of argument but not in another. Modus Ponens and the Law of Non-Contradiction, on the other hand, are invariant across fields.

The concept of field may be expanded, however, beyond the categories of Data, Warrant and Backing that Toulmin was interested in. One can add these categories, notably values, goals, morals, dispositions, preferences, and so on, or, alternatively, see the additional material as contained in the backing. The most comprehensive discussion of this is by Willard [12] where the role of values is carefully laid out. Willard emphasizes that is not merely interests or some few goals and values that may be field dependent, but entire worldviews, complete ways of imagining how everything hangs together. This in the field of Cost Benefit Analysis, everything can be converted into a cash value, while for many people not involved with CBA, such equations are at best nonsensical and at worst anathema.

So, it should be clear that I am presenting a panacea for the difficulties associated with computer - user argumentation. To the contrary, fields are not simple entities, both as a result of their internal complexity the sheer range of material they cover, and for a number of other reasons as well.. First of all, it is very important to remember that any given person belongs to a wide number of fields, which are not always consistent one with another and can, indeed, conflict. A Parent, for example, wants high quality education readily available, while the same person is also a Taxpayer and does not want education costs too high. As if that were not enough to make the situation difficult, the obvious solution, identify the hierarchical ranking of a user's fields is not as simple as it sounds. Aside from that task simply being difficult because many users are not aware of their rankings, the hierarchy in question is liable to change depending on the context and setting. Who the user is speaking with, where the user is, and the role the user is playing in a given situation will impact on the hierarchy. And the icing on the cake is that even on direct questioning a user may not be aware of his or her priorities, a well known problem facing statistical surveyors.

When humans communicate with each other they typically share a number of fields. In fact, they are, more often than not, members of the same field and communicating within its parameters. The fields they do not share will generally not

One Man's Fields

At Work	At Home
Software Designer	Husband
Wage earner	Father
Taxpayer	Taxpayer
Employee	Catholic
Co-worker	Republican
Husband	Wage earner
Father	Skier
Catholic	Employee
Republican	Co-worker
Skier	Software Designer
Etc.	Etc.

Fig. 1. The importance of an individual's fields will vary according to context and situation.

impinge, but may, of course, rear up at any moment. The point is that we most frequently argue with people we know and with whom we interact regularly. Only in relatively rare instances do we communicate in more than a superficial way with a complete and total stranger, and then when we do the initial part of the interaction focuses on finding common ground, i.e., shared fields. Frequently the very frame of the interaction, e.g., it's location or subject, can identify common fields as when one approaches an otherwise unknown salesperson. A computer, strictly speaking, has no fields and is forever interacting with people with whom it has nothing in common, and, most importantly, knows nothing about. But, given certain assumptions about why the User is interacting with the machine, and the objectives of the program, certain field commonalities may be assumed. That is, the program interface may be designed with any number of beginning assumptions about what the typical user will want, why they have begun the interaction, and what are the goals of both the User and the computer. Still, the identification of the subtleties of the User s field allegiance is a complex and intricate task.

Despite these difficulties and complexities the notion of fields and even the venerable Principle of Verifiability can be used to facilitate computer-user communications. I suggest several ways.

1. The machine can assume the existence of certain "shared" fields depending on the information context. (If the interaction concerns airplane bookings, then, e.g., the program will assume that the client is willing to board a plane, can afford a ticket, and so on.) As a result there may be tentative hypotheses made about common values.
2. Given the assumption of one or more fields, a loose version of the Principle of Verifiability can be assumed as a tool to separate facts and values using the computer's knowledgebase indexed by the field. (That is, the "facts" the computer relies upon given the presumed field sharing are taken to be shared unless interaction proves otherwise. The program assumes, e.g., that "Airplanes can fly" is a fact and not a value.)
3. The machine can inquire as to relevant field memberships not initially established and make highly friable assumptions on their basis. (The program might ask if the client has children, is employed, etc., and form one or more

queries based on values assumed to be shared by members of the identified fields.)
4. The computer can use field conflicts to determine the hierarchical ranking for an individual user. (A smoker who wants to travel but is unwilling to take a non-smoking mode of transportation is identifying a field hierarchical ranking.)

Rather than discuss these four proposals in isolation, I would like to use the time to discuss two examples in which they can come into play. The first one is a simple and familiar situation already quite common, while the second is more difficult and presents greater challenges.

In example one, which we will call the Air-Go[1] example, a user, Clara, is going to use the web based travel service, Air-Go, to book a reservation, a now familiar www experience. We will suppose that Clara lives in Boston, is attending a conference in San Francisco, wants to attach some holiday to the trip, and is on limited budget via her grant. This immediately identifies a set of fields to which Clara will *apparently* belong simply by virtue of arriving at the site and filling in the base information[2]. I call these fields the *Passive Context* as they, along with a plethora of potentially irrelevant data are passively assumed true unless otherwise contested.

Clara's Fields [Passive Context]: Air Traveler, East Coast Resident, Visitor to West coast

Obviously, this does not tell us a great deal. But Air-Go has a special feature called the Air-Go Travel Agent which engages the client in an inquiry aimed at determining the client's exact needs. In addition, and for the purposes of our example, most crucially, it inquires as to the purpose of the trip. It learns that Clara is going to a conference, but wants to spend some front end or back end time traveling. The program even asks which – front or back – is preferred, and learns that back end is preferred, but front end is acceptable. These fields now help to form the *Active Context* which can be shared more specifically between Clara and the program. Clara's fields are now extended to the following.

Clara's Fields [Active Context]: Air Traveler, East Coast Resident, Visitor to West Coast. Woman, Academic, Conference Attendee, Holidayer, Budget Conscious, Flexible Traveler.

Now the program can offer Clara a flight with dates she might not have considered, but which offer the lowest fares. In addition, Air-Go might try to tempt her with a particular hotel or spa that might meet her needs. Of course, nothing here is new. Companies frequently give questionnaires to their clients just to identify fields and interests and thereby tailor offerings to them. Or, companies track visits to web sites, check bookmarks, examine cookies, etc. to create an

[1] At time of writing there is no web site www.air-go.com, and this is used solely as an example.

[2] I say "apparently" because I do not want to confront the problems associated with misleading information, intentional or otherwise. The program can only, at this stage, assume honesty.

interest profile or stereotype of a user. In any case, the goal is to find out what fields an individual user belongs to so as to be able to make assumptions about their interests, which, four our purposes translates as beliefs, values and goals.

In the Air-Go example, there is no argumentative interaction. Clara may be offered a number of alternatives which she can accept or reject. No real attempt, at least in any site I ve visited, is made to pursue the matter. Imagine a hard sell program which inquired as to what was wrong with the offer, or which was more aggressive about tailoring to Clara s desires. I suggest we stop where we do because simply because we do not know enough about the User to be more persuasive. Perhaps this approach can be used to create a richer User Profile within the context of an interactive discursive exchange between a user and a program based on an item of contention. I.e., a computer can identify the beliefs, values and goals of a user through an interactive process and use that information to create a nuanced argumentative environment in which disagreement can be pursued. If we accept the essential idea of a field, and conclude that fields always contain within them facts, values, goals, and mores, then determining that a user is a member of the field permits the program to identify potential values. More, the program can even identify potential or actual field memberships which are likely to result in value conflicts which can then, in turn, be tested for their hierarchical status. Indeed, argument often proceeds effectively by bringing up value and goal conflicts. Being forced to choose between conflicting values indicates both a field hierarchy preference as well as something, a belief, goal or value, that must be abandoned or adjusted by the user.

A crucial difference between the commercial use of interest identification and the persuasive approach being explored is that in the argumentative context the goal of the process is to move the user from one belief or attitude to another. It is one thing to suppose that someone interested in air travel may also be interested in car rentals, another to conclude that a user interested in grade-school education will favour the release of public bond issues. In the former case the offer of rental cars is just ignored, in the latter the attempt to persuade may fall flat or even backfire.

In other attempts at dealing with computer-user argumentation [6] we have tackled a fairly thorny issue, viz., cigarette smoking or diet. In these examples an individual interacts with a computer programmed to try and change the user's attitude about a particular kind of behaviour. One problem we faced was determining the values an individual held and how those could be used to implement useful forms of argument that might lead to persuasion. I want to suggest now that the assumption that sets of values that form a nexus can be associated with membership in various fields, and, most importantly, these fields can be identified by the program through relatively simple interactions.

Imagine a program designed to persuade users to wear a seatbelt when in an automobile. A user seats himself down at a console and begins the interaction by choosing the subject of seatbelts. Now there are a multitude of arguments and thousands of reports and facts that can be brought to bear when building a case for the efficacy of seatbelts. But arguments rarely work if they are random or take a shotgun approach. Rather, they are best tailored to the needs, beliefs, values and goals of the arguer. In other words, arguments are always geared to a particular audience [9]. The key is to identify the values held by the audience and utilized those to make the case for the values being propounded, in this case using seatbelts. It literally makes no sense to begin arguing with a User before knowing the user's position construed not as a simple discursive object such as a proposition, but rather as a nexus of goals, values and beliefs. To do otherwise is like attempting to write a coherent program by composing random bits of code.

Let us say that someone, call her Ursula, approaches a machine to discuss seatbelt wearing. Given that action we may make a number of simple assumptions about her[3]. First off, we can assume that Ursula does not generally wear seatbelts, but is at least occasionally in situations where their use would be expected. Further, we know that she is at least open enough both to the issue and as a person to engage in a persuasive interaction about the subject, *regardless of how ready she is to change her mind.* Beyond that, Ursula is a cipher. More specifically, we do not know which of many possible reasons are those that persuade Ursula to disdain seatbelt use.

When, in the natural course of events, we enter into an argument, or even a discussion, with someone, we make many assumptions about their beliefs and values. In addition, we make assumptions about the way the conversation is going to proceed [7]. Typically, if someone utters something like, "I don't like to wear seatbelts," or "I don't wear seatbelts," a good arguer will first try to find out why. As a rule, one never argues with a claim, but only with the reasons for a claim [4] to do otherwise is to lay yourself open for traps. Someone, for example, might want to ban hanging, a proposition with which you agree, but because it is too quick, not exactly your position. So, it is common in real argumentation for the first question to be, "Why?" How the machine parses this response, I leave to more schooled minds, but suffice it to say that the information is best coming from the User rather than being laid out within a range of choices.

Jut as in the commercial Air-Go example, the machine makes assumptions, but it can also learn something about the user, then the possible reasons for disagreement may be narrowed down. If the gender, occupation, hobbies, and social status of a user are known, then the possible arguments being presented can often be prioritized. Stereotypically, a male motorcycle fan will not offer breast discomfort as a reason, and a mother of four small children will not likely cite the thrill of danger[4]. So a little survey profiling the user can go a long way toward eliminating any number of reasons for disagreement.

Once an initial profile is created, the program can begin to inquire what aspect of the profile is the most crucial to the position User holds. In our example, the program quickly determines that User is a woman, mother, wife, part-time

[3] I am, for the purposes of this discussion, assuming that the user is not merely curious, obstreperous, or what have you. These are certainly possible, but they do not directly address the intended purpose of the machine.

[4] Naturally, anything is possible. But determining the values inherent in fields and audiences is always liable to correction.

teacher, not religious, and a theatre enthusiast who does not smoke. By using field associations that is, by looking for connections among the values pertinent to an identified field, the program decides what arguments might be persuasive. Being a wife and being a teacher are possibilities, but not as likely as being a woman or a mother. Someone who is religious might be amenable to an *argumentum ad verecundiam*, an argument from authority, and someone who is a non-smoker may be indicating a degree of risk-aversion that is relevant to the issue at hand. The computer may have associations between woman-hood and motherhood and seatbelt wearing that are usable in this context. As a result, it can now return the query, "Does your objection to seatbelt use have to do with a] being a woman, or b] being a mother, or c] neither," and, of course, this can be put into any language form desired.

When Ursula returns the information that it is her involvement in motherhood that moves her to avoid seatbelt use, the computer can go into that field and search for associations between motherhood and seatbelts. From research on the topic the program knows there are strong value rankings associated with children, helping children, and most relevantly to this issue, *saving* children. Continuing the inquiry, it soon becomes clear that User is afraid that if she is belted in she would be unable to assist her children in the event of a collision. Ursula's position has now been identified. *Effective argumentation, whether person to person or machine to person must begin with the correct identification of positions.* Failure to determine the position held by a user easily results in arguments that are beside the point, and may lead to making assumptions about values and beliefs that are not field specific.

Once having determined that the field Mother is the most relevant and highly ranked field for Ursula, it becomes possible to work from within those field values. That is, the program can assume the values of the field being used by the client, and determine if there are avenues of persuasion that rely on those values. Still keeping things simple, the computer might offer an argument designed to persuade Ursula that she has a better chance of rescuing her children from a collision if she had been wearing a seat belt herself. In other words, the computer accepts the value hierarchy of the field, and argues that the most valued result – keeping children safe – follows from the program's desired end and not the user's. This could be supported by facts concerning survival rates, and even information regarding her use of child safety seats. If Ursula gave up smoking so as not to die and leave her children motherless, then maybe she would begin to use her seatbelt for the same reason.

There is an analogous correspondence with "facts." Facts, for the purposes of a field, are those statements commonly believed by those who subscribe to the field. Thus, facts for astrologists are quite different than facts for astronomers. Mothers, for example, will place a strong reliance on their ability to "know," i.e., intuit, that their baby is ill. Within the Mother field the intuitional liaison between mother and baby is considered very strong, and data derived from it is taken very seriously. Knowing that this mode of verification plays an important role in the field, means the program can both be prepared to receive and to react

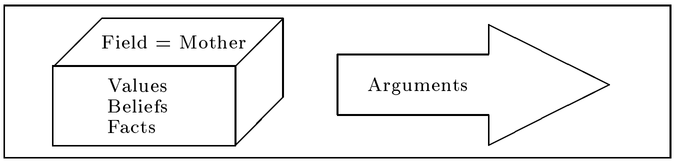

Fig. 2. When the field is identified, arguments may be drawn from within it.

to such arguments. (Vide, [4, 5].) In a sense, the principle of verification is not being abandoned, but is rather being tailored to meet the needs and values of the field.

I am suggesting here that the identification of the fields to which an individual subscribes or is committed can provide an important method for identifying the values and beliefs of an individual user. In doing that two things become possible. First, the program is able to understand the position being put forward and place it in a frame with limited relevance associations, and secondly, arguments can be geared to the values and beliefs most important to the user. This means that by paying attention to fields a computer attempting to argue with a user can play a role similar to that played by another person, that is, it can make assumptions and put forward positions based upon a shared frame of reference. In addition, by identifying the field[s] of a user, the system can sort and filter informational databases so that "facts" used and sources referred to are those that one has a reasonable expectation of being acceptable to the user.

References

1. A.J. Ayer. *Language, Truth and Logic.* London : V. Gollancz, 1946.
2. P.K. Feyerabend. *Against method.* 3rd ed. New York : Verso, 1993.
3. M.A. Gilbert. Multi-Modal Argumentation. In *Philosophy of the Social Sciences.*, June 24: 2: 159-177.
4. M.A. Gilbert. *How to Win An Argument.* 2nd ed. New York: John Wiley & Sons, 1996.
5. M.A. Gilbert. *Coalescent Argumentation.* New Jersey: Lawrence Erlbaum Associates, 1997.
6. M.A. Gilbert, F. Grasso, L. Groarke, C. Gurr and J.M. Gerlofs. The Persuasion Machine: An Exercise in Argumentation and Computational Linguistics. In C.A. Reed, T.J. Norman and D. Gabbay, editors. *Argument and Computation.* (Forthcoming).
7. H.P. Grice. Logic & Conversation. In *Studies In the Way of Words.* Harvard U.P., Cambridge, MA, 1989. (orig., 1975).
8. S.T. Kuhn. *The structure of scientific revolutions.* 3rd ed. Chicago, IL : University of Chicago Press, 1996.
9. C. Perelman and L. Olbrechts-Tyteca. *The New Rhetoric: a treatise on argumentation.* University of Notre Dame Press, Notre Dame, Indiana, 1969. (orig Fr. 1958).
10. K. Popper. *Objective Knowledge: An Evolutionary Approach.* revised ed. Oxford: Oxford University Press, 1972/1979.
11. S. Toulmin. *The Uses Of Argument.* Cambridge: Cambridge UP, 1969.[Orig. 1958.]
12. C.A. Willard. *Argumentation & the Social Grounds of Knowledge.* Tuscaloosa: University of Alabama Press, 1983.