Modes of Correspondence between Information System and World

Colin Potts

Georgia Institute of Technology, College of Computing

For an information system (IS) to be useful, it must correspond in some way to the world. This is often described as a relationship between the model of the world that the IS maintains and the 'real' world outside the IS. But this talk of models and correspondence begs these questions: what does it mean for an IS to correspond to the world, and what is the purpose of this correspondence? Since an IS is a practical artifact, it is not surprising that these two questions, the question of meaning and the question of use, are tightly related. The purpose of this paper is to elucidate the relatedness of the questions, to sketch answers to both, relate the discussion to current debate in the field of requirements engineering research, and to ground the discussion in the context of a real example. The word 'philosophy' today has acquired unfortunate connotations among some technical designers, as though it were a synonym for worthless speculation. Lest there be any doubt about practical significance of such analysis, the focus of this paper is on the type of IS, *fiduciary applications*, in which people place a lot of trust and invest a part of their personal identity, and for which the theoretical issues discussed herein are of great practical, moral and political significance.

Correspondence in Theory

Surrogacy, Correspondence and Speech Acts

As early writings on database semantics (e.g. Kent 1978) noted, surrogates in the domain of the system represent phenomena in the world. From this pragmatic perspective, therefore, correspondence between IS and world amounts to the accuracy with which surrogate entities in the IS serve as proxies for the phenomena they track. The correspondence between world and language also has a long history in philosophy. Indeed, for the twentieth-century analytic tradition, it was the primary issue, fueling debates about verification, truth and meaning. Since an IS can be construed as a linguistic system, because a database is essentially a set of propositions alleging facts about the world, it is a small step from the surrogacy account of data semantics to the correspondence theory of linguistic truth.

But as the analytic tradition loosened its hold on twentieth-century philosophy, many philosophers turned their attention to how language relates to action and practices. An important theme in these explorations was the notion of the speech act, which stemmed from Wittgenstein's (1953) language games and Austin's (1962) analysis of how to do things with words. Searle (1979) in his partial taxonomy of speech acts highlights the different modes of correspondence that may exist between words and the world, characterizing some (e.g. the fact-carrying statements that had hitherto been the exclusive focus of philosophical attention) as requiring a 'word-to-world fit', whereas others (such as commands) require a 'world-to-word' fit. In other words, the satisfaction conditions of a factual statement are that the world corresponds to the meaning of the words, whereas the satisfaction of a command requires that the world be brought into correspondence with the meaning of the words.

Problem frames and adequacy

Jackson's (2001) influential theory of problem frames is a parallel exploration of modes of correspondence between symbol systems and the world. In the case of problem frames the symbol systems are abstractions of the problems that an IS will solve rather than sentences, and the world is the problem domain or domain

of discourse of the system. Requirements engineers are encouraged to think about the requirements for the system as being adequacy constraints on the relationships within and between domains. The different problem frames involve different types of requirement and different notions of adequacy. Problem frames are a lens through which we view and judge systems and proposals for systems. We will judge whether a system 'works' or that it is 'correct' or 'right' differently according to the problem frame that we regard it as addressing.

For example, the *information-display* problem frame is the one most typically assumed in discussions of surrogacy and data models. In it, the IS is regarded as an oracle that answers questions about the world. However, it typically does so by consulting or making inferences over a model of the world, its database, rather than probing the world directly. The IS is adequate to the extent that the answers it provides are correct. If we need an independent measure of correctness, we simply look at the world and compare it with the contents of the IS. If there is a disparity, it is the world that is right (by definition) and the IS that is wrong.

There are, of course, difficult cases of IS adequacy when using this criterion. Not all databases represent the 'real' real world. For example, an online fantasy game may contain a large database storing the attributes and exploits of mythical heroes and goblins. This system may resemble some bizarre personnel tracking system, but the world that it tracks is internal to the system and totally fictitious. If a dispute arises about whether I have qualified yet to be a level three wizard, there is no real world to look at for verification. Although technical criteria would classify this system as an information display, this is not really the problem frame we should be using. Rather, managing the mythical world is a work-piece problem, and the system helps users construct and edit that work-piece. The use of a database in the implementation of the system does not change the type of problem that it solves nor the adequacy criteria we should apply. But, of course, these points are not unique to information systems; they apply to language, too. In everyday speech, we make lots of apparently factual claims about imaginary, counterfactual or dream situations, claims that clearly cannot be checked even in principle against reality.

Less exotic are systems that track the social or institutional world for which there are no real physical correlates. It may be true that the amount of money in my bank account has some relationship to where a pile of gold ingots (a very small pile) resides in a vault somewhere, but the relationship is so tenuous and mediated by such a complex web of commitments, agreements and stipulated definitions as to what shall count as monetary value, that it is impractical to check the accuracy of a banking system against the world of gold bullion. Rather, when we protest a bank statement, we appeal to a possible inconsistency between one surrogate, the bank statement, and another, an audit trail of transactions. This is a purely internal mode of correspondence checking. However, this difficulty is not unique to the IS realm; it applies equally to the meaningfulness of language about social arrangements. It was Searle (1997) again who addressed the peculiar nature of institutional reality, which he argues stringently is no less part of the 'real' world than physical phenomena, by appealing to its basis in webs of institutionalized and particular speech acts.

A second important problem frame is *commanded behavior*. Here, an IS is charged with bringing about situations in the world through the medium of commands to actuators. Viewed through the lens of commanded behavior, a system is responsible for constantly modeling the state of the real world so that it can affect the world through its commanded actions. In a variant of the commanded-behavior frame, a problem is seen as not controlling but influencing behavior through the intermediary of a human user. Being rational agents and not mechanical devices, these users decide whether to act on the system's recommendations against a background of tacit and explicit knowledge, professional and personal responsibilities, predilections, motivation, and social forces. In the case of pure command and recommendation systems, however, the job of the system is similar (or at least, the job of the system-plus-user composite system, in the latter case): this is to bring the world into correspondence with the model.

The adequacy of the system is judged accordingly: does it bring about the changes in the world that it is supposed to?

In the work-piece problem frame, the world being modeled does not 'really' exist. Rather, it is an abstract world of symbolic objects, such as documents. These the system brings into existence. Since they are non-physical, it can do this without paradox. Thus, a document that a word processor brings into being has no original existence outside the word processor. Once created, however, the document is as much a part of the real world as any other socially constructed object.

This action of a work-piece problem-frame IS of bringing an object into being has a counterpart in the study of ordinary language: the declarative speech act. When a duly authorized person in the appropriate circumstances says "I now pronounce you husband and wife" he or she is not making a true statement about a fact but is bringing that fact into existence through his or her very words. The adequacy of work-piece applications is therefore ultimately aesthetic. Just as we may judge whether marriage should be defined one way or another according to our values of what marriage is and how society should be organized, so we judge ultimately what a document is by rules of well-formedness. A word processor 'works' to the extent that the documents it produces are really documents according to those rules. It is not to be judged by asking whether the documents it produces 'really contain that text' or by demanding that it operate a printer in a particular way. To ask those questions would be to miss the point of what a document is and what word processors are for.

It matters greatly what kind of problem we judge a system as addressing. But most real systems solve many problems and sub-problems. These problems are seldom all of one type. So, a given IS has to be viewed in different ways according to the type of problem we are focusing on. Which problem frames dominate our thinking depends on our analytic purpose. This is a subjective judgment, not an objective property of the system itself.

Fiduciary Applications

Asking what kind of system one has or what kinds of problems it is addressing has consequences far beyond the methodological choices made by technical developers during requirements analysis or system testing. They become morally and politically pregnant in cases where an IS enters into a fiduciary relationship with users or other stakeholders. As we shall see in the example section, how a voting machine is construed drastically affects and is affected by how we construe the process of voting and running elections.

Of course, we place trust in all systems that we find useful to some extent. If we had to 'look over the shoulder' of every IS we used, we might as well not use them at all. By *fiduciary* applications, I am referring to those applications in which there is an inherent and explicit recognition of trust relationships and concepts in the system's problem domain. Whether an application is fiduciary is not a matter of necessary and sufficient conditions. It is an open-textured, family-resemblance concept (Wittgenstein, 1953), but three factors appear important:

1.) Users or other stakeholders can suffer greatly from system inadequacies.
2.) Users or other stakeholders invest some of their personal identity in the process that the system supports.
3.) The system can be construed as acting on behalf of the user or other stakeholders or in a capacity of trusted agent. This means that conflicting aims or goals of others should not compromise the support the system provides the user or stakeholder in question.

Fiduciary systems are those in which people have special trust. They are not necessarily financial transaction systems. They are not necessarily safety-critical (and often are not). The risks that people want

to be protected from may not always be what are regarded on purely technical grounds as security breaches or invasions of privacy.

Example: Voting and Election Machines

To illustrate the principles outlined above, we now discuss them in the context of a specific application that gives rise to numerous fiduciary requirements: electoral systems and voting machines. It is the fair conduct of elections and the universal suffrage of the citizens that legitimates the Government of a democratic society, and there are few more significant aspects of the civic infrastructure than the integrity of its democratic process. The conduct of elections provides a particularly apposite example of fiduciary requirements engineering and the perceived accuracy and adequacy of systems. When we participate in an election, we want the results to be 'correct' and we want the technology that is used to operate correctly. But what 'correct' means, what correspondence, if any, we insist on between the operation of an IS and the real world of the election depends on how we construe the process of casting a ballot. The idea of problem frames can help elucidate this.

The US Presidential Election, 2000

During the U.S. Presidential election of 2000, there was enormous publicity surrounding fiduciary requirements of the voting equipment, ballot designs and electoral process. Some of these were directly tied to the operation of electromechanical machinery, but most were procedural in nature and are likely to leave a legacy in future plans for voting machines, Internet-based voting, and election management software. Although the news media focused on the situation in Florida, because the extraordinary closeness of the vote in that state necessitated recounts and placed the outcome of the national election in doubt, news soon emerged of irregularities and procedural flaws in other states. And in Florida, it was not just the design of the ballot used in some counties and the accuracy of machine-recorded votes that raised questions. It also became clear that many voters were disenfranchised because their eligibility to vote was not recognized, and the policies for accepting and counting absentee ballots could not be consistently applied because of the ambiguity of procedures governing them. This is therefore a colorful example of a world and an IS that were not and could not be shown to correspond with the adequacy that many observers thought necessary, and it provides a warning of what the future might hold if computer-mediated Government operations become standard. However, apart from its remarkable social significance and newsworthiness, the 2000 election is in many ways typical of how systems and their requirements judged adequate or not on the basis of inexact estimates of whether the system's surrogates adequately substitute for social, organizational or normative phenomena.

Surrogate Adequacy in the Election

Three major issues of surrogate adequacy arose in Florida during the 2000 U.S. Presidential Election: (1) whether the intent of voters was adequately measured by votes recorded by the voting machines and the procedures and judgments made by precinct workers while recounting the votes; (2) whether a citizen's right to vote as a legal entitlement was adequately represented by the computer-based electoral roll obtained from other Government agencies that recorded voter registrations; (3) whether the occurrence of the election in real time and its placement of voting places adequately represented election policies and laws.

Votes in the World and Votes in the Machine

Voters in several Florida counties complained during and after the election that the ballot was difficult to understand and the voting machines operations and its recoding of votes were not transparent. It has been alleged that many voters who intended to vote for Vice President Gore were confused by the ballot design and instead voted for a fringe-party candidate. Other ballots were rejected because the vote-counting equipment registered no vote having been cast. The voting machines in question punched holes on the ballot form where the voter indicated. Several thousand ballots, however, showed pressure marks or evidence of partial punching but no holes. Huge controversy ensued during the month that the election result was in doubt and in the aftermath of the election, concerning these spoiled ballots and their dimpled and hanging 'chads', and I do not intend to enter that debate here. For present purposes, the central issues concern the fiduciary requirements for the technology: (1) What is a vote that we can know that one has been cast deliberately and according to a voter's intent? and (2) How can the integrity of the vote be assured, or if there is doubt or dispute about the outcome, how may it be validated, so that the public can trust the announced outcome?

Regarding the first question, we could adopt a definition by fiat that a vote is whatever is recorded on the ballot and registered by a correctly working voting machine. So, if I intend to vote for Laurel but erroneously indicate by pencil mark, lever pull, or button push that I am voting for Hardy, then I am by definition voting for Hardy. Or this strategy could be refined somewhat by defining a vote as whatever vote was most recently recorded before a "commit" operation was selected. The two strategies are essentially the same: We can finesse the problem of surrogate adequacy by defining it away. There is no world, only IS. There is no vote other than the pencil mark, electronic record, or dimpled chad. It is these that we count to obtain the result of the election, and it is these that are real. Everything else is in the voter's head and is therefore outside the boundary of our concern.

Another strategy is to implement the system identically, but recognize that a vote is a well-defined intention of the voter and that the recorded mark, symbol or chad is but a surrogate the adequacy of which is always open to reinterpretation. According to this view, the reality outside the system is accessible and not merely a mental construct. If we need to challenge a count, it is possible to examine the surrogates that allegedly represent votes cast and scrutinize them again to determine what the voter's intent was. This, it can be argued, is something that only a person can do, not a voting machine.

Clearly, both solution strategies are unsatisfying. At its worst, the first arrogantly denies the problem, essentially by stating that if the map and the territory disagree, you must believe the map because really there is no territory. But the reason why we have elections is to determine the will of the people not to determine how many holes or indentations occur on a set of specially sanctioned pieces of card. The holes are a surrogate for the will of the people.

The second solution strategy is equally problematic, but its abuses are different. At its worst, this solution implies that if we do not like the territory, we can always change it by redrawing the map. In the case of many systems, it may be possible to go back to the source for reconfirming information, and the reasons for accepting or rejecting this information may be subject to independent evaluation. Thus, if my citizenship application turns on when I first entered the country, and a system records a date that I contest, it may be possible to assemble evidence from other sources that will convince the agency responsible for the system to change the record. Even then, what is really happening is an informal dialogue in which I ague that your surrogate should be replaced by a better one. Without going back in time, I cannot actually present you with the phenomenon of my entering the country. Thus the decision whether to change the surrogate value (i.e. whether to decree the change a "correction") is itself a judgment that is warranted by the judged reliability and integrity of the new evidence. If the authorities refuse to accept that the IS is wrong and insists that my evidence is incorrect, there is little that I can do. In the case of the election, it was

impractical to identify the voters whose ballots were in question, would have been illegal to ask them their intent after the election even if they had been identifiable, and their subsequent answers to such a question were it to have been asked may not have reflected their intention on the day of polling because of changes of opinion, errors, or tactical voting decisions.

There is no perfect solution to the problem, but when much depends on the adequacy of surrogates, as it does in the case of an election, it is surely important to emphasize the problem during requirements analysis and the early design phases. Identifying items of data whose provenance and reliability may need to be computed and stored in addition to their values must be done at the earliest stage. In the absence of such deliberately applied heuristic guidelines, we are left with the possibility that claims for the reliability of the system (meaning here, its accuracy) will go uncontested.

Denial of Service

Strict rules usually govern the duration of polling on the day of elections. Other rules govern the validity of absentee ballots based on their source, date of vote and date of arrival, and their provenance. In the U.S. Presidential Election of 2000, many polling stations were reported to have closed on schedule with voters still waiting to vote. Others were claimed to have stayed open illegally to let waiting voters vote. There are many possible reasons for the presence of voters at closing time, including voters waiting until the last minute to vote, delays caused by equipment malfunctions, understaffing and the increased time it took for voters to cast their ballots in cases where the ballot was long and the voters confused by it. What is again important for our purposes here is not to cast blame or comment on the political consequences, but rather to emphasize that the real-world phenomena of "polling station" and "election day" are themselves constructed and negotiable entities and that their surrogates in a more automated system may have very different properties. These surrogates are resources used in the voting process, and the delays that occurred naturally in the 2000 election are instances of a type of problem that makes computer-based systems with similar functions vulnerable to denial-of-service attacks that would be unfeasible in the case of many distributed physical locations.

In the case of in-person voting, only votes cast in a polling station on the day of election count as votes. The corresponding rules governing absentee ballots are less easy to define precisely, and yet it is these that have to be formalized in the case of support for remote online voting. In an experiment conducted by the U.S. Department of Defense in 2000, members of the armed services and civilians working for the Department of Defense and stationed overseas were able to cast votes using computers. The computers so equipped and the software running on them therefore could be taken to be surrogate polling stations. Alternatively, they could be taken to be an electronic version of the kitchen table and mailbox used for hand-marked absentee ballots. This is not merely an issue of which metaphor is the most appropriate: Federal Law prohibits polling being conducted on Federal property (presumably to avoid the perception of coercion), and yet the workstations used were necessarily at secure locations on military bases. Issues of what counts as a polling place, the duration of the election, and a valid vote can only become more problematic as the site of voting moves, as many advocate, to public libraries and even the home. Not only are there issues of surrogate adequacy, the trustworthiness of a distributed election system will depend on how confident the public and officials can be that it is immune to denial-of-service attacks.

Another way in which physical locations mask the need for surrogates in a computer-mediated system is institutional role that is played by counties and states in the U.S. system of government. (Again, the details are specific to the U.S., but transcend U.S. Government and elections in general.) A person may only vote in the county in which he or she is registered, and the ballot is unique to that county. In addition to President of the United States, the same ballot may also contain entries for Congressional representatives, state representatives, judges, or local school board. By having voters come to a physical

location, these choices can be packaged on a single ballot form. This is not an option for absentee voters who used the experimental DOD system. Counties have to coordinate their elections through a central administrative agency that can therefore deliver the appropriate ballot to the absentee voter. It has to handle or at least mediate the eligibility validation process, and then once the votes are cast it must deliver the ballot to the appropriate jurisdiction. If the institutions of 'polling station' and 'election day' did not exist, and all voting were to occur online at any time during the weeks before an election deadline, there would be no need for a single ballot. You could vote for president one day and local dog-catcher the next, as long as you only voted once for each. The issue addressed in the previous section about when the vote has been cast then becomes doubly problematic: Could a voter change a vote at any time before the election was to close?

Counting As A Voter

In the U.S., citizens may register to vote in several ways, including when they apply for a driver's license. The assembly of the electoral roll therefore depends on integrating information from more than one Government agency and IS. It is not uncommon for people to arrive at a polling station only to learn either that they are not on the roll at all or that they have been registered elsewhere and must go there to vote. It is also possible to be told incorrectly that you have already voted and may not vote again. In recent years, criticism of the registration process and the inadvertent disenfranchisement of citizens through bureaucratic mistakes have led to calls for conditional voting in which a voter whose eligibility is challenged may nevertheless cast a ballot that is marked as being conditional on the person's eligibility being validated. This, of course, requires that the ballot or electronic surrogate be linked in some way to the identity of the voter, thus undermining the privacy of the ballot.

Discussion: Problem Frames for the Voting Process

We can get a better perspective on the role of surrogate identification and specification in the elaboration of fiduciary requirements if we explicitly enumerate the problem frames [Jackson, 2001] or metaphors [Potts, 2001] employed when we talk about the problem domain. Consider the following stages of the election process and the problem frames/metaphors that operate at each stage (Table 1). Which fiduciary requirements and how many of them we take seriously depends on which phases of the election process are seen as within the problem boundary and which problem frames or metaphors dominate our thinking about them. For example, registering to vote is temporally disconnected from the voting process and may involve different organizations and systems, but it could be construed as falling within the domain of discourse. If so, we are forced to consider the accuracy of the roll from the first. It is not an "input" to our system, but a set of surrogates that our system is responsible for.

Similarly, we could regard voting as a kind of writing or articulation of intent. If so, then the ballot is a form that the voter "writes" on, and the dominant problem frame that governs our thinking is one of work-pieces. If the voter is a writer and the voting machine an amanuensis, then it would seem much more appropriate for election officials to take the responsibility subsequently of seeking to understand the voter's intent. This is what we do with ambiguous documents. If, on the other hand, the marks on the ballot are physical objects that count as surrogates for a person's vote, then they act as records of facts. It is more natural in this case to consider voting a kind of "data entry" and the voting equipment as part of a composite information-display machine. In this case, issues of provenance become more important because a raw data record contains nothing that guarantees its valid origin. Should some votes have to be recounted, or some votes be challenged, it is important to be able to individuate the surrogates corresponding to those voting events.

Table 1: Events in the domain of elections, classified according to the dominant problem frames and resulting issues of surrogate adequacy.

Domain event	Dominant problem frame / metaphor	Surrogate adequacy issues
Citizen registers for voting	Command behavior / Delegate (Citizen as machine operator)	Is the voter identified accurately? Is the entitlement created?
Voter presents himself or herself as a voter	Information display / Oracle (Voter and election official as questioners)	Is the voter present in the roll? In case of conditional vote, can ballot be associated with voter?
Make decision	Information display / Oracle (Voter as questioner)	Are the correct candidates presented? (Not far-fetched in case of absentee voting).
Register, undo or commit vote	Commanded behavior / Delegate (Voter as machine operator),	Is the vote recorded as intended by the voter?
	Work-piece / Amanuensis (Voter as writer)	Is the ballot document formed as indicated by voter?
	Information display / Oracle (Election official as assertor of fact; voter as its subject)	Does surrogate for voter indicate ineligibility to vote again in this election?
Count votes	Transformation / translation (Election official as clerk)	Are all ballots counted correctly (including conditional ballots)
	Information presentation / oracle (Public and media as quesioners)	Is election result as determined by number of votes (unlikely to be an issue).

If an election is seen primarily as a process of articulation in which the will of the people is made manifest, an election support system can be thought of as a kind of giant bulletin board. The most important fiduciary requirements then become equality of opportunity to express one's will (i.e. the adequacy of the electoral roll and the availability of the voting "places" at the appropriate time) and the anonymity of one's vote. Focusing on the ballot itself as a thing, however, gives a completely different perspective on what an election is. From that point of view, an election is the secure production, warehousing and distribution (i.e. counting and tabulating) of ballots. In that case, the most significant fiduciary requirement is the security of ballots. None should be lost, only high-quality ballots (i.e. those that are eligible and unambiguous) are to count, and it is these, not the nebulous will of the people that the election is about. Finally, if the ballot is not seen as a physical object but as a document, the problem of conducting an election amounts to the secure editing and subsequent analysis of forms, and the fiduciary requirements in question now become the correct reading and interpretation of marks.

None of these metaphors are necessarily better than any other. However, how the domain is construed, what problem frames dominate the thinking of the developers and stakeholders, and the problem boundary legitimated by stakeholders all affect which fiduciary requirements are the most salient during the requirements and early design processes. These considerations, in turn, have a major effect on the perceived trustworthiness of the system once deployed.

References

Austin. J.L. *How to Do Things with Words*. 1962. 2nd ed. Harvard UP, 1975.

Jackson, Michael. *Problem Frames: Analyzing and Structuring Software Development Problems*. ACM Press, 2001.

Kent, W. *Data and Reality*, 1978, 1st Books edition, 1998.

Potts, Colin 'Metaphors of Intent'. *Proc. RE'01: International Symposium on Requirements Engineering*, Toronto, August 2001: 31-39.

Searle, John R. *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge UP, 1979.

Searle, John R. The Construction of Social Reality. Free Press, 1997.

Wittgenstein, Ludwig. *Philosophical Investigations*, 1953. 2nd ed. Basil Blackwell, 1958.