

Knowledge technologies and the semantic web

Kieron O'Hara and Nigel Shadbolt¹
University of Southampton

While the Office of Science and Technology commissioned this review, the views are those of the authors, are independent of Government and do not constitute Government policy.

1 INTRODUCTION

In this paper, we discuss cyber trust issues relevant to knowledge technologies and the new area of development, the semantic web (Berners-Lee et al. 2001). Knowledge technologies are technologies enabled by recent technological developments that allow much more intelligent machine engagement with the documents, services and other objects on the World Wide Web. They manipulate and create knowledge, that is, usable information. The major fault lines for trust management are in making sure that the input to knowledge manipulation processes is trustworthy and in ensuring that the processes themselves are trustworthy. Their limits and margins for error must be known and predictable.

There are many ways of creating or maintaining trust in this domain. We set out a number of approaches or 'tactics for trust'. These include: allowing scrutiny; maintaining transparency; transferring ownership from experts to stakeholders; exploiting the minimal transitivity of trust; and requiring evidence of identity, provenance and certification. Above and beyond these, we suggest restricting interactions to a small set of agents; using formal methods to make data less 'scruffy'; using 'calculi of trust' to determine when and when not to place trust; using new technologies to allow interrogation of and dialogue with agents; replicating computational processes; and using knowledge technologies to manage knowledge more effectively.

These tactics need to be combined in active trust management strategies. It is important to maintain agile policies for managing trust, including the collection of rich sets of metadata about knowledge sources and agents and ontologies for expressing trust requirements. It is essential to maintain the distinction between trust and trustworthiness, so that signalling trustworthiness does not become detached from trustworthiness itself. It is necessary to ensure that functionality is not sacrificed to trustworthiness. And, finally, privacy has to be sufficiently protected so as not to undermine trust.

The discussion in this paper leads to suggestions for many opportunities for interdisciplinary research collaboration. These include determining the extent of the distinction between online and offline trust; examining whether a utilitarian or a moral notion of online trust is appropriate; understanding the nature of trustworthy knowledge acquisition; investigating why personalized interactions are perceived as more trustworthy; examining how brands and reputations contribute to trust; and considering how effective procedures for the maintenance of knowledge bases can be developed.

The semantic web, conceived as an extension to the World Wide Web, is a potentially large area for Internet development. The web is, of course, massive, containing in 2002 about 2.5 billion fixed documents – 550 billion documents in all, when databases and other information sources that users can access via web forms are added into the total, between them holding 7.5 million gigabytes of data (O'Hara 2002a). The web is now so large that the various navigation tools, including hyperlinks and search engines, are coming under increasing strain. This is leading to pressure to move to the semantic web, which, for the eXtensible Markup Language (XML), the Resource Description Framework (RDF) and ontologies, allows systems many more inferential possibilities than the World Wide Web, whose underlying language, the Hypertext Markup Language (HTML), is more concerned with looks than content.

There is no way of knowing whether the semantic web will take off as the web did; much will depend on whether the tools and formalisms devised are as successful at selling the concept to the heterogeneous band of users that would be essential to its large-scale take up. In 2004, the

semantic web had not decisively left its home in the computer science department. However, this paper will remain relevant because many of the claims within it will apply to the web as well and because, however the web is upgraded, its sheer size dictates that various functionalities, such as intelligent search and the customization of content, will inevitably be developed in some form.

We focus our analysis in this paper on likely developments over a five to ten year horizon. This is a very uncertain field and the technology could develop in a number of ways. However, we focus on the likely development of the web and the semantic web, given the current direction of technology and the existing set of problems on the web.

This paper is structured as follows. We begin by scoping our enquiry. Section 2 sets out our assumptions about trust. This is not intended to be an in-depth contribution to philosophy or sociology; neither is it intended to undercut other discussions of trust that appear in this volume. Rather, the aim is to set out our preliminary understanding of the concept of trust that underlies our argument in this paper. Section 3 briefly sets out the technological background of the semantic web, and Section 4 introduces the concept of knowledge technologies.

The creation of trust can happen at many levels; individual technologies can attempt to be trusted or trustworthy, or it may be that groups of people and organizations set up institutions for the wider purveyance of trust. Accordingly, our discussion of trust treats these levels separately. Section 5 looks at the possible ways of promoting trust that individual designers might be able to use; we call these the *tactics* of trust-creation. We, therefore, claim that the issues surrounding trust at a wider level are relevant to *strategies* of trust-creation (strategies that might involve the artful deployment and combination of more than one of the tactics of trust). We discuss these in Section 6. Bringing strategies into the discussion implicitly suggests that many aspects of cyber trust stretch beyond the abilities of individual designers to address them and that, therefore, cyber trust is a problem that a number of disciplines can throw light upon. The interdisciplinary aspects of the problems and likely solutions are discussed in Section 7.

2 TRUST

We begin with a brief overview of our understanding of the concept of trust. As we focus on the important semantic web notions of services and technologies, we will be looking at ways of ensuring that inputs to and outputs from such services and technologies are trustworthy. We make the assumption, common to many commentators, that there are strong analogies between online and offline trust (Corritore et al. 2003).

We give less attention to trust as a security issue. Clearly trust has an important security dimension, but secure infrastructures are not covered in this paper (see Jones & Randall review at www.foresight.gov). The key issue for the semantic web, as with the World Wide Web, is that services deliver reliable output, which may be as much a social issue as a security one. The semantic web and knowledge technologies do not seem to raise any particular security issues peculiar to themselves (the security issues pertaining to e-commerce, largely with respect to transactions and transfers of monies, are the central security issues for the web, see Camp (2000)).

In general, information downloaded from the web tends to be fairly reliable. Given the number of people posting information, this in itself is a minor miracle. The World Wide Web is a major source of information for many people – academics obviously, but also for various more or less naïve users who download often introductory information about, for example, their medical, financial or legal problems. While the majority of this information is relatively reliable, impartial or explicit about its partiality, web users are generally trusting of it. The trick will be to keep this information reliable in future generations. The issue is much more a social than a security one.

To define terms, trust is at minimum a binary relation, between someone or something that does the trusting, and someone or something that is trusted. We will call the actor that trusts the *principal*, and the actor that is trusted the *agent*. In this paper we will make no distinction between principals or agents that are human and those that are 'artificial'. We will equally make no distinction between principals or agents that are individuals and those that are institutions. In other words, we assume that the human notion of trust is usefully exportable, if only metaphorically, into these non-human realms. This is a controversial assumption, argued against by Hardin (1999), for example. A defence of the assumption for practical purposes can be found in O'Hara (2004a) and

Corritore et al. (2003).

Furthermore, the trust of a principal in an agent is generally context-bounded, so the agent is trusted to *do* something; this something may be open-ended, or it may be quite specific (Grandison and Sloman 2000). This extra context is not essential to trust, because trust can be across the board. But, in World Wide Web and semantic web contexts, in general, there is some contextual limitation of the scope of trust. In the terms employed by Corritore et al (2003), World Wide Web and semantic web contexts generally produce *specific* trust as opposed to *general* trust.

2.1 Functions of Trust

Much discussion of trust begins with a functional definition.² In a survey of empirical research on cyber trust, it has been noted that conceptualizations of cyber trust are often conflicting, which has militated against comparisons and the synthesis of results (Grabner-Kräuter and Kaluscha 2003). The multi-dimensional and, indeed, second-order nature of the trust construct is clearly a problem. O'Hara (2004a) discusses some of the difficulties of defining trust and some of the anomalies that crop up when trust is defined rigidly. He argues that a better methodological approach is to define the sphere of life or society in which the author/audience is interested and then to try to trace all the various trust-like phenomena that impinge upon that sphere. Most areas will be affected, not only by trust, but by other phenomena such as risk, expectation or inclusion into moral communities that are very hard to disentangle from trust. It is, therefore, unsurprising that different conceptions of trust should produce a range of functional characterizations (Misztal 1996).

The most common approach with respect to cyber trust is the definition of trust as an epistemological construct that reduces complexity and/or transaction costs in conditions of uncertainty, generally following Niklas Luhmann (1979). Interactions with other actors are rendered simpler because the trusting actor simply foregoes the necessity of investigating other actors' credentials or of putting in place systems of performance metrics, invigilation regimes or management structures. In an uncertain situation, the trusting actor simply behaves as if there is more certainty than there actually is and, therefore, is absolved from having to perform all the extra tasks that uncertainty would have loaded upon it.

This is an important function of trust and we follow the majority of commentators in making this assumption (Grabner-Kräuter and Kaluscha 2003). In a context where many of the interactions are either commercially-based (e-commerce), or where the aim is to create a cooperative environment, for example, creating and acting on a plan involving a number of autonomous agents, then trust is likely to be a central complexity-reduction strategy.

However, particularly where the Internet is concerned, there are other functions of trust that deserve consideration. In particular, we might also follow Emile Durkheim (1893) and Talcott Parsons (1949) in looking at the role of trust in projecting a consensus of values and thereby helping to integrate a community. In communities with a moral dimension, which are held together by a set of values, trust often plays the role of signalling inclusion into that moral community. Much empirical survey evidence from offline contexts seems to suggest that trust is an inclusive, value-based capacity (Uslaner 2002). In such contexts then, trust, far from being the effect of extended displays of trustworthy behaviour, acts as a cause of trustworthiness (O'Hara 2004a).

This is extremely relevant to the Internet as a whole. Thanks in part to its development in academic circles, the Internet is a very value-laden space and the extension of the Internet, via the web, to non-academic contexts, including political and commercial contexts, has caused a substantial cultural clash – where so-called 'newbies' are looked down upon by the 'founding fathers'. In general, the governing values of the Internet among its creators are anarchistic, based on liberty and free expression. Such values have helped the technical development of the Internet quite dramatically (for example, with the open source software movement), but equally have made the (so far successful) grafting of a commercial marketplace onto the Internet somewhat fraught.

In many ways, these two interpretations of the notion of trust pull against each other, and ways of extending the first notion may prevent the extension of the second, and vice versa (Lessig 1999). Nevertheless, each function of trust should, as far as possible, be respected in a value-laden space such as the Internet and – depending on its course of development – the semantic web.

2.2 Properties of Trust

Trust is often best characterized by its functions. However, it tends to be recognizable by its properties. As trust is generally desirable, the exploitation of these properties to achieve trust would, if feasible, be a very useful development. However, trust, being a second-order phenomenon, is rarely as reproducible as that.

One tactic is to devise axioms that reproduce some of trust's properties, particularly its limited transitivity (see Section 5.7). It has been argued that trust is not transitive (Povey 1999) – it certainly is not completely so, but it, nevertheless, exhibits some transitive properties, even if these are unpredictable and often unintentional (Grandison and Sloman 2000). Basically these properties are dependent on the main unobserved property of trust, which is that it builds up slowly and is dissipated quickly. In general, an agent must prove that it is trustworthy by meeting expectations (which may be moral or merely practical) within understood parameters.

The extent of the trust, the number of tasks for which the agent is trusted, will also be limited – the agent is trusted to do *something*, not trusted unrestrictedly. In other words, trust is usually specific to a particular task, or function or promoted service. If an actor is trusted, then this is usually shorthand for the actor's being trusted for a specific implicit task or set of tasks. Such models of trust may help in the development of automated systems for trust management and trust creation (Grandison and Sloman 2000).

2.2.1 The non-specificity of trust

Unfortunately, the triggering condition for trust is highly non-specific. There are no answers to such questions as:

- How many times does the agent have to achieve the task before it is trusted?
- Under what range of circumstances does the agent have to achieve the task?
- What set of interests should characterize the agent?

The last question, about interests, is noteworthy. If it is in the agent's own interests – independent of the principal's interests, and of the agent's interests to be trusted – to perform the task, then it may be that trusting the agent to perform the task would be premature (Hardin 1999). What must happen is that the agent's interests must somehow become aligned with those of the principal. This may happen with the explicit alteration of the agent's interests, for example, by the principal paying it a fee. Or the agent might be altruistic. Or it might be that the interests of the agent can be manipulated by the principal. Whatever the case, the performance of the task is in the principal's interests – this is why it has to trust the agent – and in order for the principal to trust the agent, it has to be confident that the agent's interests and its own are suitably aligned.

As Corritore et al. (2003) argue, trust can come in various shapes with very different properties – which is another way of saying that the properties of trust are highly unpredictable or that modelling trust in general may well be inappropriate or indeed impossible. For example, Corritore et al. draw a distinction between slow trust, built up over time (and therefore relatively rationally based), and quick trust, which occurs when relationships are created and then cease to exist within a short timeframe (for example, when a team is assembled to perform a specific task). They also distinguish between cognitive trust, where the principal has good rational reasons for trusting, and emotional trust, where it does not. These distinctions can be helpful, but they may simply reduce to the proposition that trust can be rational or not and that it may build up quickly or slowly.

2.2.2 The withdrawal of trust

The loss of trust is more straightforward. Failure to perform the task, in apparently appropriate circumstances, will in general result in the forfeiture of trust by the agent in that respect (and possibly in other respects too). This is not to say that trust will necessarily be lost, but that it is liable to be lost. It is also possible that a suitable explanation of the failure from the agent, or an explanation of why the circumstances were inappropriate for trust, will save the situation.

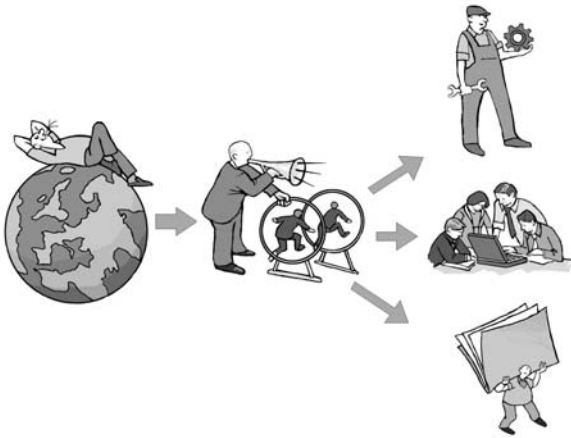
Trust, though, need not be all or nothing (Grandison and Sloman 2000); it may be associated with a level, for example, a real number between 0 and 1. In that case, withdrawal of trust may mean a

decline in the value assigned to this parameter, but not necessarily to 0.

Principals must have knowledge of the agents they are trusting. This is one example where trust, when it operates smoothly, takes on transitive properties; trust transfers through intermediaries between the actors at each end of the chain. This clearly does not always happen; hence, our caveat that trust is only partly or occasionally transitive.

The performance of many tasks is organized by brokers or project managers. For instance, a principal A might trust an agent B with the performance of some task, while B's contribution is to decompose the task into subtasks and to assign the subtasks to trusted agents of B's own (towards which B now acts as principal). A knows nothing of the other agents. The situation is as depicted in Figure 1. If one of B's agents, for example C, fails to perform the task, then A does not cease to trust C, of whose existence it may not even be aware. It ceases to trust B; the untrustworthiness of B's agents transfers back to B.

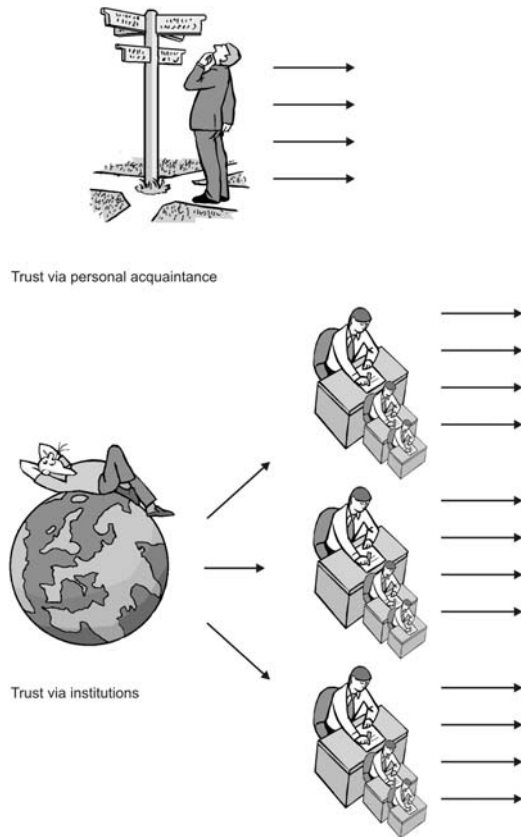
Figure 1 Proxy agents



Indeed, many of the mechanisms that are used to promote trust do so at the cost of massively increasing risk. For example, consider the use of institutions to spread trust. In such a circumstance, an institution takes on the role of certifying trustworthiness and, therefore, the principal has access to many more agents in whom it can trust. The institution has to have power and sanction over those it certifies. There is still a primitive trust arrangement, but now the principal trusts the *institution* to certify responsibly. This division of labour, illustrated in Figure 2, is held to be vital for the spread of trust through even moderately complex societies (Fukuyama 1995).

However, as has been argued elsewhere (O'Hara 2004a), such an arrangement also dramatically increases systemic risk. Where a principal trusts an agent through personal acquaintance, a betrayal results in the withdrawal of trust from that agent. However, where a principal trusts an agent as a result of a certificate from an institution, a betrayal might well result in the principal withdrawing trust from all agents certified by that institution (in the absence of a further reason to preserve trust in any individual cases).

Figure 2 Spreading trust via institutions



Trust via institutions has been called *global* trust and trust via personal acquaintance *local* trust (O'Hara 2004a).

2.2.3 Trust and contracts

Trust is in many respects similar to contracts. If A trusts B, then B claims that it will bring about some state of affairs in A's interests; A then investigates no further, but will behave as if B's claims were true (for example, it will plan for the future on that basis). If A has a contract with B saying the same thing, then it will equally behave as if B's claims were true. In neither case does A have any proof that B will carry out its claims; in neither case does A have any power to bring the state of affairs about itself. So the similarity lies in the fact that A has no power to bring some state of affairs about, yet is able to plan for a world in which that state of affairs obtains. With a contract, of course, A has recourse to law should B renege, which will allow A to impose some sanctions (though not necessarily to bring the state of affairs about). In the case of trust, the only sanction against B's reneging is for A to cease to trust B, and to publicize the reneging.

However, there are two important dissimilarities with contracts that should be noted. First, trust is much more flexible than contracts. A contract usually specifies background conditions that must obtain for the contract to be valid. An allowable defence against breach of contract is that the contract was not valid at the time of the breach (independently of intentions).

Second, contracts can crowd trust out. Where contracts are binding, recent empirical studies have shown that trust relations tend not to evolve. Where a certification approach (see Section 5.4) involves tightly determined sanctions and powers for the certification authority, then trust may be inhibited (Spiekermann et al. 2001). Trust, being risky, tends to evolve only when there are no other mechanisms for addressing the inherent uncertainty in the situation. Non-binding contracts, on the other hand, tend not to inhibit the development of trust (Malhotra and Murnighan 2002).

2.2.4 Trust and knowledge

A related point concerns knowledge. Trust is a rational strategy under uncertainty. One obvious way of increasing trust and reducing risk is to reduce the uncertainty, that is, to perform some investigation of the agent. If we rely too heavily on this strategy, however, then trust relationships will not develop. They will not need to develop if the uncertainty and risk are reduced to minimal levels. Under such circumstances, the investigation would have used up a lot of work and demanded navigation through highly complex situations; as the need for trust decreased, then so would the benefits from trusting have become unavailable. Removing uncertainty altogether takes away risk, but at the potentially high cost of the investigation.

2.2.5 Trust and risk

There is an interesting, yet under specified relationship between trust and risk – in general, they have a roughly inverse relationship. People are less inclined to trust when risk is high, and so this should be taken into account. For instance, people risk money and resources in e-commerce applications and so may be less inclined to trust even though the level of security is usually higher than in other applications. There has been little work linking trust management and risk management (Grandison and Sloman 2000).

2.2.6 Distrust

The topic of distrust should be mentioned briefly, although it is not addressed further in this paper. Most treatments of trust implicitly include distrust, which, consequently, need not be treated separately. However, it should be noted that distrust is not simply related to trust. In particular, it would be wrong to assume that distrust is analyzable as trust in a complementary piece of behaviour. If I distrust you to pay me back, it is not the case that I trust you not to pay me back. In particular, Grandison and Sloman (2000) must be wrong to define distrust as 'the lack of firm belief in the competence of an entity to act dependably, securely and reliably within a specified context' (which is the complement of their definition of trust). This definition makes no distinction between, say, a randomly-chosen person whom I do not know and, therefore, have no firm beliefs about at all, and a known betrayer about whom I harbour firm beliefs about his or her untrustworthiness. I distrust the latter, but it seems correct to say I neither trust nor distrust the former.

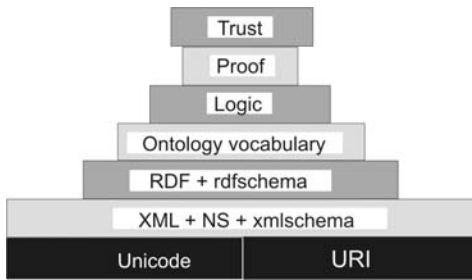
3 THE SEMANTIC WEB

The semantic web is the potential successor to the World Wide Web. As the web has increased in size, the mechanisms for navigating through it are coming under increasing strain. The move to the semantic web is expected to enable more intelligent navigation and customization of content. For those unfamiliar with the principles underlying the semantic web, an appendix at the end of this paper provides a short introduction.

Trust has always been envisaged as a key factor in the structure of the semantic web. Figure 3, developed by Berners-Lee, shows how the various layers of the semantic web are supported. Each layer 'makes sense' of the layers below. XML tells the computer what the Unicode data encoding and the Universal Resource Identifiers (URIs) refer to; RDF tells it how they are related; ontologies give a meaningful context for these types and relations, creating a holistic 'web' of meaning (see Quine and Ullian 1970). There would be no point in having this language without logic to make inferences; no point in making the inferences without a proof theory to ensure that the inferences are valid; and no point in producing the proofs without trust in the system as a whole.

With these formalisms and structures, the semantic web provides an environment in which a computer can behave intelligently; compare this to the web, where the system of HTML plus web addresses allows the illusion that the whole web is lodged on the current machine. But the data remain uninterpreted. Trust is not required, because the machine is inactive.

Figure 3 The layered view of the semantic web



Source: adapted from Koivunen and Miller (2002, p. 34).

There are two important points with respect to the semantic web and trust that we need to note at this stage. First, like the web, the semantic web is designed for anyone to operate in; the semantic web should be a space in which political activity, commercial activity and leisure activity can coexist with the inevitable scientific activity. Again, as with the web, the expectation has to be that semantic web users are heterogeneous and that the data in the semantic web are 'scruffy'.

The emphasis therefore needs to be on formalisms and mechanisms that can cope with such 'scruffiness'. For example, ontologies probably need to be partial, flexible and interoperable, rather than highly principled and fixed. But this scruffiness may also make trust a little more complicated to produce. It may be – and the evidence for this is small and equivocal – that there is an epistemological version of Gresham's Law, that bad knowledge will drive out good. For example, if datasets are going to be large and relatively 'noisy', then the inferences drawn from them will be correspondingly riskier. Risk parameters will be computable, but equally these will tend to be disregarded, especially outside scientific or otherwise statistically literate communities. As a result, the chances of information being seriously misleading may well be relatively high.

The second point concerns the current circumstances of the semantic web, which is at a fairly early stage. It is an emerging technology, and one that is intended to develop a high profile. There may be a trade-off between developing technologies that exploit the expressivity of the semantic web, and those that are provably trustworthy. Getting useful – and usable – structures, formalisms and tools off the ground, now may be more important than fostering trust across any kind of representative cross-section of the heterogeneous group of potential semantic web users. This is not to say that such a trade-off is inevitable, only that, for a short period of time at least, it may be a sensible strategy for semantic web development.

In fact, the incidence of errors may actually affect trust. Corritore et al. (2003) have a prominent role for credibility and error-proneness in their model of online trust.³ It may be that smoothly functioning systems or easy to navigate websites are necessary if not sufficient conditions for trust. Having technologies that work may well be an essential priority even for those who value trust highly. Professionalism is also valued by users; Corritore et al. cite more studies that show that cues that have an impact on user perceptions of trustworthiness include ease of navigation, good use of visual design elements, professional images of products, freedom from grammatical errors or typos, a professional look to the website, ease of search and ease of carrying out transactions. Indeed, website design is a quite major determinant of trustworthiness for lay users as compared to experts who look at information quality (Stanford et al. 2002).

4 KNOWLEDGE TECHNOLOGIES

The semantic web provides a very helpful context for the development of knowledge technologies, that is, those technologies whose aim is to transform information into knowledge, or *usable information* (O'Hara 2002a). Knowledge technologies are technologies that transform information so that it can be an input to some problem-solving process. Typical of such technologies are those that enable significance to be extracted from stores of information; these include ontologies, systems for enabling or automating the production of annotations or metadata for information sources or systems for information extraction from natural language.

Another important class of knowledge technologies is those that enable the transfer of information around an organization, providing knowledge sharing services for instance, enabling the identification and distribution of knowledge. These might include discussion spaces for documents or communication systems for remote or asynchronous meetings.

Following the account of the UK's Advanced Knowledge Technologies (AKT) project,⁴ we assume a (rough) knowledge lifecycle including the following stages: Acquisition, Modelling, Retrieval, Reuse, Publishing and Maintenance.

A knowledge technology would be expected to contribute to the manipulation of knowledge in one or more of these stages. Trust impinges on such technologies in two ways. First of all, given the scruffiness of the worlds in which such technologies flourish, one problem is how to ensure that knowledge technologies operate on trustworthy data. Second, given that the input data are trustworthy, how can we trust the processes underlying the knowledge technologies themselves, in order that we can trust the ultimate output?

4.1 Epistemological Functions of Technology

The trend, particularly since the economic downturn in 2001, is for organizations to focus less on technology installation and on reconfiguring themselves, and to switch the focus to services (Economist 2003). Knowledge technologies and the semantic web are very much in this mould. When the service ethos is combined with the epistemological advances that knowledge technologies and the semantic web make possible, the result is *knowledge services*. Knowledge services provide knowledge-based analyses through the medium of the web, for example, using ontologies for data capture and integration or for user modelling for customizing content for the reader. Metalevel brokering services are also interesting in this context; such services could provide and integrate a suite of services required for a particular organizational context.

An alternative scenario for knowledge technologies, though rooted also in organizational sense-making and consensus-creation, would be the development of technologies for enhancing collaborative work for workers dispersed remotely through space and time. Various intelligent aids to communication could, for example, integrate multimedia representations of meetings, combining movies with PowerPoint and supporting facilities such as querying with something very close to natural language, given sufficiently rich, probably ontology-driven, annotation of the material.

4.2 Trust in Content and Input

In many cases, knowledge technologies will be extrapolating structures from vast quantities of heterogeneous data with a great number of sources (for example, where knowledge services are extracting information from webpage text and generalizing over that information). In such cases, where the range of information sources is messy, the knowledge technologies must trust in the reliability of the majority of the sources so that any misleading or out-of-date data are swamped by reliable data and appear only as noise. For instance, if a system is taking information from the web pages of a department's professors to form a picture of the department's publication records, then the extent to which the system must ensure that the pages are up to date will depend crucially on the extent to which the system has to present the truth. Are there sanctions that will apply if the system produces an inaccuracy? If there are no sanctions, are there, nevertheless, opportunities (for example, funding opportunities) that could be lost with inaccurate information? Or is the aim of collecting the information merely a general provision of useful material that needs to be broadly trustworthy, but not super-accurate?

Where knowledge technologies are taking data from a single source or type of source, then that source has to be trustworthy. In order to trust that source, the knowledge technology may need to gather some information to reduce uncertainty; this information, the metadata of that source, should provide partial indications of the reliability of the source. The provenance of the source may be of a particular type (its address may be from the '.edu' range), or the source may make reference to other trusted sources. Of course, the metadata and annotations themselves must be trustworthy, needing maintenance and curation as much as the top level information.

Branding of sites is important. For example, taking information about academic sites from a central government site, such as that of a research funding body, is better if that site has a good reputation

for accurate collation of information. Such a reputation may well be encapsulated in a brand. Of course, symbolic shortcuts such as brands will be of much more relevance to the human intelligences guiding the machines than to the machines themselves. But the notion of creating an analogue of branding is intriguing. Correct identification of sites is important with any development of the brand model.

Where knowledge technologies are intended to facilitate knowledge sharing or discussion, there is an obvious problem of trust with respect to the person with whom one is sharing or discussing knowledge. If a novice persisted in firing off uninformed questions in a document discussion space, for example, much of the value of that space might be lost as a result of the necessity to reinvent the wheel for the purposes of the discussion, or the other interlocutors might lose patience with the whole process.

4.3 Trust in Process and Output

In terms of process and output, knowledge technologies must have reliable methods. Such methods as knowledge acquisition techniques, modelling languages and so on must do what they advertise. But evaluation and testing of such methods is non-trivial (Shadbolt et al. 1999). For example, many knowledge technology methods are designed to work over the web scale. If the methods are not reliable, then the output of knowledge technologies will be correspondingly less trustworthy. For trust – which obviously depends on knowledge and publicity – such evaluations must be publicized and certified (or inspectable). For instance, a knowledge modelling method may need to provide links to academic papers describing the method and its results.

Similarly, where a knowledge technology manipulates knowledge (for example, when it retrieves an answer from a repository, or when it repackages some knowledge for bespoke publication), then an important issue of trust is whether that manipulation preserves the important properties of the knowledge. So, for instance, if information is presented in personalized web pages, two different users must trust the knowledge technology to be presenting the same knowledge, even if in different forms (different languages, different levels of abstraction, etc). Good ontologies are important here; yet many ontologies may be generated automatically from combinations of smaller ones. There may be all sorts of problems with respect to issues like the integrity of referring expressions (Alani et al. 2002).

Explanation is very important. Knowledge technologies must be able to reduce the uncertainty of the users about their output. They must be open to interrogation, or provide explanation of their output, together with other useful metadata such as margins of error. This entails the development of, for example, query languages to frame an appropriate interrogation.

One important aspect of trust in this context is privacy. Much privacy is preserved by the clear separation of sources of knowledge; placing two knowledge sources together can be invasive of someone's privacy by allowing a much greater range of inference about that person. For example, the inferences available about someone on the basis of their tax records combined with their bank account history massively outweigh those possible without cross-referring those two sources.

A related concern is that of the secondary use of data. One can hand over information for some purpose – to a funding body, to a law-enforcement organization – and then an important privacy question concerns what legitimately can be done with that data. The aim of many knowledge technologies, for example, those that find and repackage information from the web, is precisely to use data for secondary purposes.

There is an interesting distinction between the American and European models of privacy. The American approach is rights-based and people are given rights to disclose or not to disclose information in various circumstances. Once information has been freely given away or published for some purpose by an individual, he or she has relatively few rights to prevent its secondary use. Europe has a data protection model where secondary use is regulated and restricted to so-called 'fair use' (that is, private data can only be used for the purpose for which the individual has given consent). Tailoring a knowledge technology to one or the other of these regulatory models may be non-trivial.

4.4 Agents

Much of the discussion on trust in the context of the web, the semantic web and knowledge technologies is complicated by the fact that many agents – and indeed some principals – are *artificial*. When trust is taken functionally, as it almost always is in the literature, as a strategy for complexity reduction, then there may seem to be little difficulty in adapting what began as a human trait to the realm of the artificial. After all, the serious problem of bootstrapping – the key issue, for example, discussed in Hobbes's *Leviathan* – can be finessed much more easily with artificial agents than with recalcitrant humans.

However, though this is undoubtedly an advantage, humans are of course able to apply a great deal of flexibility to their trusting decisions and to balance their interests very finely (though equally their reasoning may be more bias-prone). In a network of heterogeneous agents, such as an online market might consist of, agents' decisions to trust other agents will depend on effective models of trust. Such models are a matter for experiment as to what best balances flexibility and an acceptable level of risk.

Furthermore, agent environments are inherently distributed and decentralized. Such potential security ideas as central databases listing security clearances are not going to be possible on large scales. Security must be achieved on the basis solely of an accurate evaluation of an agent's own credentials. And the dynamic nature of agent environments means that security policy will have to be dynamic too.

Artificial agents being dynamic and built for interaction, it may well be that virtual organizations are created and disbanded on the fly for particular tasks. Such organizations of agents may only be in existence for relatively short periods of time. Questions of trust will almost certainly arise about how such a mayfly organization can take responsibility for decisions or output, and how the user may place trust in it (see Section 5.4).

Humans will also have to learn to trust their artificial agents. Again, in many ways this issue is continuous with the issue of any principal trusting its agent. However, trusting an artificial agent involves being *au fait* with the properties of that agent and its suitability for the virtual environment or market in which it operates. For example, if the agent is a negotiating agent, the principal needs to know how advantageous the negotiation algorithm is and how suited it is to the market.

The extent of trust will of necessity vary, depending on the powers transferred from the principal to the agent. Variations on this theme include: What decision rights has the principal transferred to the agent? Can the agent gain access to the principal's resources? And, is the agent acting as principal to other agents?

5 TACTICS FOR TRUST

We move on now to a review of approaches to trust. First, in this section, we look at a number of basic approaches to the creation and sustenance of trust, what we have called *tactics* for trust. Then in Section 6, we look at how these tactics might be combined and deployed; in other words, we look at potential *strategies* for trust.

5.1 Transparency

The first tactic for trust that we consider is that of transparency. Here, the tactic is for the agent to open up its activities to the scrutiny of the principal. If some black box processes are governed by the agent, then the agent can allow the principal access to the workings of the processes, opening up the black box.

The principal would not be assumed to be constantly observing these newly transparent processes. If the principal performed such rigorous oversight, then most if not all the gains of its trust strategy – complexity reduction, lowered transaction and information-processing costs – would be lost. The incentive for the agent to behave in a trustworthy way is that the principal could, at any time, check any of the newly transparent processes. Transparency is generally assumed to be a key strategy for trust.

There is some empirical evidence for this. For example, trust in automatic recommender systems

can be increased by a conversational interface and the disclosure of the recommender's personalized user model (Corritore et al. 2003).

There are problems with the use of transparency to promote trust, however. First, much depends on the complexity of the agent's services, and how likely the agent calculates it to be that the principal will discover any underhand practices. Second, and perhaps more importantly, is the effect that discovery of untrustworthy behaviour might have relative to the expectations of the principal. The problem here would be if the agent was using relatively unreliable sub-processes. For example, it might be employing heuristics or subcontracting work to a risky agent. This problem might also be compounded if the agent, in contact with subagents of its own, had access to the principal's resources.

It may be that the unreliability of these sub-processes would be unlikely, except in the extreme, to affect the successful outcome of the process as a whole. But if the expectations of the principal were unrealistically high, then it may withdraw trust from the agent. Transparency, in conditions of uncertainty, will always contain the risk of loss of trust as the flip side to its benefits.

5.2 Transfers of Ownership

A second tactic is to transfer ownership of certain processes to the principal. The idea is that stakeholders take responsibility for processes or artefacts, rather than allowing experts or authority figures to adopt that responsibility themselves. An example would be where a community might take over ownership of knowledge-based resources, such as ontologies, from the knowledge engineers who might otherwise be able to impose those resources without significant input. The development of such resources might well be done properly, but the air of mystery might be off-putting for the stakeholders (Domingue et al. 2001).

The problem is that developing knowledge-based resources is a difficult and time-consuming job. The key is to create usable tools that allow resource development to take place in the absence, or relatively low profile, of a knowledge engineering expert. For example, Tennison's Adaptive Presentation Environment for Collaborative Knowledge Structuring (APECKS) tool allows collaborative asynchronous discussion and construction of ontologies (Tennison et al. 2002).

Another example is the use of open hypermedia to control associative linking from web pages, as, for example, in the Conceptual Open Hypermedia Services Environment (COHSE) environment (Carr et al. 2001). The use of embedded hyperlinks to simulate associative linking has always been a relatively odd strategy, as it is the author of the page that controls the associations. It is as if Proust's narrator had bitten into the Madeleine and found himself remembering, not his time at Combray with Aunt Leonie, but rather some episode from the baker's past (O'Hara 2004b). It has been argued that taking control of associative linking could increase trust, for example, by allowing the principal access to alternative views not necessarily endorsed by the web page's author (Sunstein 2001; O'Hara 2002b).

A final example is to reduce the automation of processing, so that the principal does more of the processing itself. This involves a step back from the usual knowledge technology approach, which is to automate tedious though knowledge-based tasks. If we consider what this idea would involve we can imagine a search engine looking for web pages. At present, the engine would be expected to do some of the processing (as, for example, with Google's PageRank system, which orders the hits in a cogent way) and the principal does some more (as with Google, it is the user who looks through the hits and discards those that are of no use). The aim of knowledge technologies is to shift the balance towards automatic processing and away from the principal. If the balance goes too far the other way, towards processing by the principal, then the whole principle of knowledge technologies will be undermined, since all the knowledgeable processing will have been exported away from the technology. Indeed, this points to the downside to this tactic. If stakeholders take ownership of processes or resources, this means they take responsibility. This might transfer some risk to them, or it might commit them to providing resources of their own. In turn this may lead to a lack of enthusiasm, and it may be that fewer such resources are built successfully. As we have noted, the semantic web is at an early stage of development and this is one example where trust might profitably be sacrificed in the short-term in order to secure the development of the semantic web into a medium with positive, visible and quantifiable benefits.

5.3 Exploiting Transitivity

Where trust already exists, a third tactic is to exploit transitivity of trust when it occurs; recall that trust is not genuinely transitive, though often takes on the properties of transitivity. That is, if A trusts B, and B vouches for C, this in itself provides a reason for A to trust C (not necessarily a decisive reason – hence, trust is not genuinely transitive). A related tactic, which we will not treat separately in this section (although much of the discussion implicitly applies), is to exploit the *distributivity* of trust. If A trusts a group X, then that may give it a reason to trust the Xs in X, though again this will be hedged around with many caveats. For example, what will A trust the Xs to do? The same thing in each case? Or particular subtasks of the global task? Distributivity is a more complex topic, and we focus here on transitivity.

There are advantages and disadvantages to this tactic. An advantage is that it does not rely on the creation of institutions. The Internet is remarkably institution-free, which is one reason why trust is so difficult to manage online (O'Hara 2004a). Despite the fact that the Internet is the global technology *par excellence*, trust online is actually very often local. Transitivity is rooted in direct personal acquaintance, and so a system that is based on transitivity can be grafted onto the local trust networks that one would expect to find online.

A related point is that there are remarkably few sanctions available to online institutions and so institutions are relatively unable to police their charges. An institutional approach will always be at something of a disadvantage online, although infrastructure changes may alter this (Lessig 1999). There is also some empirical evidence that institutional approaches, such as seals of approval or kitemarks, are not particularly exploited by users, at least in e-commerce contexts (Corritore et al. 2003). It may be intriguing to investigate the reasons for such a shying away from institutional approaches. One reason, in the context of European Union regulation, which was discovered from empirical analyses, is that mention of European Union regulation seems to produce a false sense of security (Spiekermann et al. 2001).

A disadvantage of exploiting transitivity is that it is inevitably limited. Beyond more than a few links in the chain, the connection between principals, the Parsonian community of interests which those principals may feel that they have in common and which underpins their trust relations, will be very stretched indeed. In relatively small 'worlds', such as individual academic communities, this may not be too much of a problem. Certain prominent figures, either famous people or branded sites, might also ease the problem of the ebbing in transitive trust if they appear somewhere along a chain. If B (trusted by A) tells A that C (unknown to A) endorses X, then A may not be altogether inclined to trust X. But if B tells A that Tim Berners-Lee endorses X, the chain may not degrade quite as rapidly. Another disadvantage is that such transitive chains do depend on reasonably robust networks of trust being around in the first place; they do not address any bootstrapping problem.

A third disadvantage is that identity becomes essential as the chains of trust become very long; there must be some kind of certification procedure, otherwise the chains will degrade. This will require the inclusion of institutions and the consequent possibility of sanction, and so may only be possible in fairly specialized contexts. The Internet is generally opposed in ethos to too much of an institutionalized approach, and so systems of, for instance, digital signatures, though useful in many contexts, may meet resistance from many Net users (Lessig 1999). There is a definite and uncomfortable tradeoff here; one of the charms of the Net is that on it 'no-one knows you're a dog', but, equally, it is useful sometimes to be able to prove you are not a dog.

Identity is also a problem in very short chains of trust, down to the limiting case of two actors, and ensuring identity is extremely important. The well-known issues of identity theft and dealing with multiple identities are as important with knowledge technologies and for the semantic web as with other areas of online interaction. The point we are making, however, is that as the chains get longer, it becomes increasingly difficult for a principal to manage its interactions (as it may then be going through a number of intermediaries, only one of which needs to be unreliable to scupper the entire interaction). In a world where chains of trust do get very long – and the semantic web is envisaged to be such a world – the institutional trade-off will appear to be more pressing.

James Hendler's 'web of trust' initiative (Golbeck et al. 2003) aims to exploit the transitivity of trust to create a social network that can extend beyond the immediate personal acquaintances of its

members. They provide metrics for inferring trust at remote parts of the chain and use such algorithmic methods for generating metadata for authentication. This approach has an interesting take on the bootstrapping problem, as it is intended to determine trust from annotations to documents that are not explicitly relevant to trust. However, a system of digital signatures is also required to authenticate entries. Various other network-based schemes are possible, for example, as described by Richardson et al. (2003), or exploiting techniques for analyzing networks determined by ontologies (Alani et al. 2003a).

The Friend-of-a-Friend (FOAF) system is an RDF schema that allows users to develop an interlinked set of statements about agents to build a web of acquaintances.⁵ The identity problem is finessed by linking the descriptions with emails; again, a system of digital signatures is also required.

Yolanda Gil and colleagues (Gil and Ratnakar 2002) have suggested that an annotation system called TRELIS⁶ could be used to bolster trust. Documents and other information resources can be annotated by users, providing assessments about argumentative structures, the quality of the arguments within those resources and relations to other resources. Such assessments go towards the ability to suggest measures of credibility that are entered explicitly by individual users and then summarized (averaged) and presented to new users. This is another way of exploiting transitivity as the users will accept these credibility assessments if, and only if, they trust previous users. It is also a good example of the sort of exploitation that can be made of the semantic web.

Certification of identity is an important adjunct to the tactic of exploiting transitivity. We discuss this aspect more explicitly in the next section.

5.4 Provenance and Certification

Identity, provenance and certification together make up a big element of the likely direction of online trust management. The provenance of some information is the history of the item insofar as it can be traced, together with information about its originator (which will bring us to the problem of identity).

One example of the provenance problems that the semantic web will throw up can already be seen in the world of virtual organizations and also in grid applications. The semantic web is likely to see virtual organizations form dynamically, very possibly coming in and going out of existence in extremely short periods of time. The trust problem here is how a user can trust information released by such short-lived dynamic coalitions, given that data will be modified depending on what services the virtual organization performs. In other words, how can the user determine what process generated the resulting data when the virtual organization that produced the result may have ceased to exist?

Provenance may be seen as an annotation recording how some data were derived, showing how data were passed between services and altered as a result. The user may then be able to step through the process via the services. Some architecture must be in place to facilitate the recording of these decisions and actions, and this architecture should support reasoning or navigation through the workflow of a virtual organization, as well as storage of the annotations, that is, the provenance data. To achieve this some centralized architecture will be required to oversee authentication and non-repudiation (Szomszor and Moreau 2003).

Centralized architectures are also important for certification of identity. We have already seen, as for example with Hendler's 'web of trust', that certifying identity is crucial for the operation of many ideas. Note that with a certification approach, the certifier does not necessarily vouch for the agent's trustworthiness; the certifier merely certifies the agent's *identity* – from which the principal must deduce whether the agent should be allowed access to resources. Most certification approaches involve a two-step process: first the key is bound to the agent by the certifier; second the principal gives access rights to the agent, independently of the certifier.⁷

For example, in an open agent-based environment, as we have noted, agents must interact with agents whose identities they will need to verify, as they will not have come into contact with them before. Such agents, particularly when they control some of their principals' resources, and particularly in negotiating or commercial contexts, will have to decide whether to act on requests and how to assess assertions. This will require examination of other agents' credentials. Such

credentials will include properties of the agent, such as membership of accredited organizations, age or host of the agent and recommendations from and delegations by other agents. In particular, delegations and recommendations (estimates of trust values) are passed about between agents in communities as ways of signalling their beliefs about the trustworthiness of third party agents.

All this must be able to be performed dynamically. Tim Finin and colleagues in Maryland (Kagal et al. 2002) have developed a dynamic framework, based in the semantic web language DAML+OIL (DARPA Agent Markup Language + Ontology Inference Layer) and a standard agent framework (such as developed by the Foundation for Intelligent Physical Agents (FIPA)), which provides authorization and credibility assessments. The security framework is based very strongly on digital signatures, generated using Public Key Infrastructure (PKI).

The advantage of this approach for an agent environment is that it is distributed. The agents can be held accountable for their actions because they have to sign all queries and requests with their personal key. This removes the need for a central database of security clearance which would not be appropriate in the decentralized agent world. Instead, an agent platform, associated with a security policy, is able to check an unknown agent's credentials whenever it tries to register.

So, simple verification of an agent's credentials works as follows. All the credentials that are relevant to the inquiry are presented at the time of request.

In order to use its services, a requesting agent must send all required credentials along with the request for service. The service agent will check its knowledge base, and question other agents about their beliefs in order to verify the credentials. Suppose agent A has an alarm service which requires that requesters be AAAI members. The security policy of agent A also states that the agent XYZ should be trusted to verify AAAI certificates. An agent B sends A a request to use the service along with its certificate from the AAAI CA. This certificate states that the bearer of this certificate is a member of AAAI. Agent A asks agent XYZ to verify the certificate. If the certificate is valid then agent B is authorized to use the alarm service. If agent B did not send the required certificate or sent an invalid certificate, its request would be denied (Kagal et al. 2002, p. 30).

An example of this is an agent-oriented PKI system described by Hu (2001). In this system, there are two types of certificate: identity certificates for humans and their agents; and authorization certificates, which represent authorizations by entities, including the public key for the granting entity, the public key of the entity that receives the authorization, the authorization itself and so on.

It should be noted that any trust policy that involves using certification systems or digital signatures, such as PKI, does not solve the trust problem, but it shifts it one step along. The user has to trust the certification system and the institutions that check the signatures; this localizes the trust issue and simplifies trust management – and is thereby extremely valuable in itself – but it does not remove the problem.

In this context, Marianne Winslett and colleagues' system TrustBuilder is an interesting approach. This exploits trust negotiation, where agents – strangers to each other – iteratively disclose credentials until sufficient trust has been built up to secure the transaction required. This has the happy effect of helping secure trust across domain boundaries so that agents subscribing to different trust systems can establish trust between themselves (Winslett et al. 2002; Grandison and Sloman 2000).

This is how such a negotiation might work.

A service agent A only allows employees of XYZ Pvt. Ltd. to access its services, and accepts delegations from these employees. Agent B approaches agent A with a credential from AAAI. Agent A decides that the credential is not good enough and asks the agent B to prove that it is an employee of XYZ or if B has a delegation from an employee. Agent B possesses a delegation from Bob who is an employee of XYZ and sends this delegation to A. A verifies the delegation and the chain of delegations and decides to authorize agent B's request (Kagal et al. 2002, pp. 30-31).

Thought also needs to be given to how to manage certification obsolescence. It may be that certification systems used for trust management become obsolete or compromised, or are merely superseded by other systems. Or, in the case of networks of trusting agents, they may simply lose their critical mass and cease to be effective. Principals who are exploiting certification systems need to have strategies for dealing with such obsolescence and managing the transition to new certification systems. This is particularly true when certified agents are of crucial importance in safety- or mission-critical applications, and where their *de facto* decertification might cause the whole application to crash. There is an intuitive distinction between certificate obsolescence and certificate revocation, but, equally, it may be hard for a principal to make that distinction.

5.5 Alternative Tactics of Restriction

So far we have been describing the tactics of restriction. The section on exploiting transitivity (5.5.3) explored the possibilities of restricting interactions with agents you 'don't know', while the section on provenance and certification (5.5.4) suggested restricting interactions with those who do not have the correct certification. There are other ways of restricting interaction with classes of agents that are possibly untrustworthy.

For example, the Advanced Knowledge Technologies project runs a semantic web application called CS AKTive Space (Shadbolt et al. 2003). This application takes information from a number of heterogeneous and distributed resources describing the state of the computer science discipline in the UK, continuously harvesting and screenscraping content, and maintaining a central store of information. The resulting snapshot of the discipline, mediated via an ontology, is then presented to users through various visualizations.

The reliability of this picture depends on the quality of the information gathered and the maintenance of the datastore. If information is gathered from an unreliable source, then the result would be untrustworthy information that might pollute the whole dataset. The solution pursued by AKT is to restrict attention to branded websites, that is, web sites that are presumed explicitly to be trustworthy, such as the web sites of relevant university departments (using the '.ac.uk' suffix as a brand), or the information gathered by research councils such as the Engineering and Physical Sciences Research Council (EPSRC). By using brands in this way, the domain of information can be usefully restricted. The development and monitoring of brands as vehicles of trust is extremely important (O'Hara 2004a).

Another example of a possible restriction would be to confine oneself to one's own research (or another) community. Communities of practice are seen as essential knowledge managing entities (Wenger 1998). They are vital for training new practitioners and for maintaining corporate memories. As such, there has been a good deal of work on identifying communities – often reaching beyond one particular organization (Alani et al. 2003a; McDermott 1999). Communities identified by relatively mechanical methods – such as Ontocopi, which identifies communities based on analysis of domain ontologies (Alani et al. 2003a) – could be another basis for restriction of interaction, being more meaningful than anything exploiting transitivity and yet more open than a system based on brands.

5.6 Formality

A leading characteristic of the semantic web, and indeed the web in general, is the 'scruffiness' of the data and the heterogeneity of the users. Trust could be extremely problematic, simply on the grounds that anomalies may be extremely hard to spot. Policing the semantic web, therefore, may be difficult. Formal methods could be of some interest in the context of the semantic web. They are, without doubt, somewhat contrary to its spirit. Nevertheless, they might serve to restrict interactions to those that can be modelled and, therefore, are relatively predictable. They can be used to lead users away from the scruffier bits of the online world and towards more straightforward bits.

There has been a relatively small quantity of work in this field. However, the general approach has much in common with the subject of the next section, the development of 'calculi of trust'. Whereas the formal approaches we envisage are designed to model the domain, calculi of trust are designed to model the trusting behaviour itself.

5.7 Calculi of Trust

Calculi of trust are attempts to model trusting behaviour and, as such, have a particular relevance in the software agent world, where trusting behaviours have to be produced entirely artificially. However, it should be noted that primitive logical frameworks rarely have the expressive power to model the typically complex reasoning required to maintain trust relationships (Grandison and Sloman 2000).

Dimitrakos and Bicarregui (2001), for example, have set up the basic elements of a model in axiomatic form. They use multimodal and subjective logics to derive inferences. These axioms

were drawn up in the particular context of e-services, with the hope that formal reasoning will facilitate formation of trust relationships and resolution of conflicts.

The model is interestingly complex and sensitive. The authors do not assume that trust or distrust are necessarily transitive. They have interesting methods for extending the discussion to consider the trusting of intermediaries; different relationships between the principal, the intermediary and the intermediary's agents (for example, does the principal even know of the existence of the intermediary's agents?) lead to very different trusting behaviour, and axiomatic analysis can expose interesting types of behaviour and isolate the possible types of intermediary.

For instance, Dimitrakos and Bicarregui suggest there are four types of intermediary. A *transparent* intermediary identifies the agents to the principal. A *translucent* intermediary identifies the existence of the agents to the principal, but not their identities (for example, telling the principal that its goods will be sent by courier without specifying which courier). An *overcast* intermediary hides the existence of the agents. A *proxy* intermediary actually acts as the agent itself, but does not reveal its own existence (that is, it acts in another's name without disabusing the principal of the assumption that the other is actually operating).

These categorizations then lead to various axioms (Dimitrakos and Bicarregui 2001). Examples include:

1. Trust (distrust) is not transferred along an overcast intermediary.
2. Trust is transitively transferred through transparent intermediaries.
3. Trust (distrust) in all subcontractors of a transparent intermediary is transferred to an inclination to trust (distrust) the intermediary.
4. Trust is transferred anonymously through translucent intermediaries.
5. Trust in an adviser is transferred to the recommended parties.
6. Distrust in recommended parties is transferred to an inclination to distrust the adviser.

Golbeck et al. (2003) use a method to compute trust. They create graphs of networks of trusting agents, developed using the FOAF schema (see Section 5.3). The method then computes measures of trust by looking at the strength of connections between the two relevant nodes (agents) on the graph. They are able to calculate a maximum and a minimum amount of trust between two agents; they also create a weighted average giving a recommendation for trusting. The algorithm for computing this searches the graph for paths between the two agents, taking the value of a direct edge connection as the trust value where such an edge exists and, where it does not, recursively determining a value of trust relationships between all the neighbours on a path. The algorithm ensures that no agent is trusted by a principal more than another agent closer to the principal on the graph.⁸ The trick with any formal calculus of trust is to capture the richness of behaviour in the face of uncertainty.

5.8 Interrogation and Dialogue

Interrogation, conversation and dialogue have long been seen as keys to trust. Testimony on its own is rarely seen as a sufficient guide to controversial action; the author of the testimony must be cross-examined in order to establish his or her *bona fides*. The argument goes back to Plato, who gives Socrates the following passage in the *Phaedrus*.

SOCRATES: You know, Phaedrus, writing shares a strange feature with painting. The offspring of painting stand there as if they are alive, but if anyone asks them anything, they remain most solemnly silent. The same is true of written words. You'd think they were speaking as if they had some understanding, but if you question anything that has been said because you want to learn more, it continues to signify just that very same thing forever. When it has once been written down, every discourse roams about everywhere, reaching indiscriminately those with understanding no less than those who have no business with it, and it doesn't know to whom it should speak and to whom it should not. And when it is faulted and attacked unfairly, it always needs its father's support; alone, it can neither defend itself nor come to its own support (*Phaedrus*, 275de).

In other words, the written – because of the properties of written script – is not trustworthy. It cannot be restricted to those who should see it. It cannot be personalized to the reader. It cannot add anything, it cannot enlighten the reader any further. And it cannot argue, or enter into a dialectic (O'Hara 2004a).

This sort of argument is still adhered to. In law courts, for example, a signed affidavit (that is, a written account of some event) is only permissible in a relatively small number of circumstances, such as when the content of the testimony is uncontroversial. Otherwise, witness testimony has to be given orally, and the witness must engage in a conversation with a properly accredited representative of the defendant. And so specialized has this conversational task become, that years of training are required, and those people (barristers) who possess the expertise can command huge salaries. Some even become media personalities – and all because there is a general presumption in society that Socrates' argument above is valid. Similarly, PhD candidates are examined orally, committee meetings depend on face-to-face discussion rather than on the submission of written argument that the chairperson can evaluate, and so on.

This is not to say that the written word is inherently untrustworthy; only that the spoken mode of communication includes many features that enhance trust. Interestingly, many of the developments of the semantic web allow precisely the sorts of dialogue and personalization of content that provide these features (O'Hara 2004b). In this sense, the semantic web actually blurs – as have earlier technologies – the boundaries between the spoken and the written. Dynamic content retrieval and publication reproduce the instantaneity of the spoken, and the responsiveness to an interlocutor. And, as such, they allow certain types of interrogation that may well extend the boundaries of what we trust. Such interrogation allows interlocutors to rate the value of information the other provides, which is an important determinant of trust (Corritore et al. 2003).

For instance, D3E (Buckingham Shum and Sumner 2001) is a document discussion environment that allows structured discussion of a webpage or document. This is a discussion that even the author can take part in. Artequakt provides personalized content for readers, synthesized dynamically from web documents – in its case, biographies of artists (Alani et al. 2003b).

Another interesting example of blurring the spoken/written distinction can be found with the open hypermedia system – Conceptual Open Hypermedia Services Environment (COHSE) (Carr et al. 2001). COHSE allows the reader, rather than the author, to generate associations. Transferring control of associative linking from the author to the reader is a prime example of a technology of the written allowing the incorporation of some characteristic properties of the spoken (O'Hara 2004b).

5.9 Redundancy

Replicating processes and introducing redundancy is a further way to help promote trust. If some process is deemed risky, having several different agents run the process, and then having a relatively transparent process of arbitration between them is advantageous. Replication tactics work particularly well in a distributed or grid environment where different nodes of the grid will have different computational capabilities, nodes will not always be available and network connectivity may be relatively unreliable. In such a case, replicating computation on multiple nodes can improve performance and increase trust by ensuring that no one (possibly flaky) node will bear the full weight of responsibility for some piece of processing (Li and Mascagni 2003).

The severity of the downsides to a redundancy tactic will depend on the type of inference being replicated. First, there are obviously the excess costs introduced by replication, since the same processing has to be done more than once. However, if the processing is relatively straightforward and costs are low, then the tactic may be useful. On the model of a computational grid, for example, the paradigm is for powerful low-cost computation to come on stream as a result of the grid arrangement, in which case the costs of redundancy may well be outweighed by the benefits.

Second, there are areas where some intelligence or creativity might be exploited in reasoning. In this event, replicating processing might pull the net result to the centre, ruling out riskier, but possibly more creative solutions. On the other hand, if the processes are relatively straightforward, or if the creation of trust is more important than extracting value out of processes, then again the replication tactic might prove to be valuable.

5.10 Other Forms of Knowledge Management and Presentation

Finally, we must not forget other, perhaps more mundane, ways that knowledge technologies can increase trust and trustworthiness. Recall the knowledge lifecycle (Section 4). The final knowledge

challenge in that cycle is that of maintaining knowledge bases. Knowledge maintenance, which is greatly aided by knowledge technologies, will aid the process of keeping knowledge bases up to date and correct, and will therefore improve performance of that knowledge base.

For example, methods to model knowledge in order to anticipate what knowledge needs to be junked would be useful. Methods for determining inconsistencies are also of value. In short, merely ensuring that knowledge bases are kept trim will increase trust not only in them, but also in those knowledge technologies that use such knowledge bases as input.

Another knowledge challenge is that of publishing. Clever methods of presenting personalized knowledge to a user in a way that allows its significance to be gauged should also increase that user's trust in the knowledge. For instance, Gil and Ratnakar's (2002) TRELIS system presents arguments about a document's veracity in a structured way. Buckingham Shum's D3E (Buckingham Shum and Sumner 2001) also allows a structured argument to develop about a text, which allows the reader to follow whether the document stands up to scrutiny. Indeed, in general, information content provides useful cues for users to assess trust, and provision of content that is appropriate and useful for the audience is very important for establishing trust (Shelat and Egger 2002; Corritore et al. 2003). Semantic web technologies to personalize content will clearly be of great importance here.

Table 1 Tactics for creating or sustaining trust

Tactic	Description	Costs
Transparency	Allow principal access to hitherto closed processes, black boxes	Potentially open to creating mistrust, if expectations are too high
Transfers of ownership	Allow stakeholders decision rights & responsibilities	Stakeholders may be more reluctant to put in effort than an agent
Exploiting transitivity of trust	Where a trust network already exists, extend it via transitive (or, on occasion, distributive) extensions	Neither transitivity nor distributivity are perfect models of trust. Plus this strategy cannot address any bootstrapping problem
Certification	Create some institutional support for digital signatures, thereby securing provenance	Institutional structures are contrary to the anarchistic value ethos of the net, and thereby might work to reduce trust (cf Durkheim). Doesn't address bootstrapping, as the principal still has to trust the certification system and authorities
Restriction	Increase trust by policies designed to avoid interaction with the non-trustworthy.	May be arbitrary. May be over-limiting. Hard to evaluate the efficacy of the tactic.
Formal methods	Use formal methods to avoid dealing with the scruffier parts of the web.	High modelling overhead. Plus the whole development of the web, with its heterogeneous users, has encouraged scruffiness. Many of the richer parts of the web are scruffy.
Calculi of trust	Use formal characterizations of trust relationships to govern when an agent should trust.	Trust, being a second-order phenomenon, is hard to model successfully. Such a system is likely to lack the flexibility inherent in trust.
Interrogation	Submit documents, web pages, etc, to interrogation and scrutiny.	Technology in the early stages.
Redundancy	Run processes several times in parallel.	Extra computing costs. For certain types of process may be opportunity costs, producing 'safe' average results rather than risky, creative ones.
Knowledge management	Use tools for knowledge management to maintain knowledge bases and keep them accurate, up to date and trustworthy.	High maintenance overheads.

5.11 Summary

We have sketched out a number of tactics for establishing and/or extending trust and these are summarised in Table 1.

In the next section, we turn to the strategies for trust.

6 STRATEGIES FOR TRUST

Trust management will continue to be very important on the World Wide Web. Developing strategies for trust, which might involve deploying some of the tactics mentioned in Section 5 in intelligent ways, dynamically configuring various tactics depending on the context of the transaction, will be essential if trust is to be maintained as the Internet user community increases.

6.1 Trust Management

Policies governing trust and trust management are essential. In particular, security requires a wide understanding of the domain in question, together with specifications of how actions ramify (Schneier 2003), and should be high level, abstract and integrable with heterogeneous applications and platforms (Grandison and Sloman 2000). Indeed, attention needs to be paid not only to the relationships between the principal and the remote systems with which it communicates, but also other components in the system such as the interactions over underlying systems, for example, communications services (Grandison and Sloman 2000, Schneier 2003).

Such management models must be alert to the difference between a system designed to block some accidental incident and a system designed to block malicious action. The former should, if well-designed, completely eradicate the type of accident it is meant to counter, although it could well have unintended consequences further along the line. The latter, however, is of necessity at risk from further malicious attempts to get around it.

We should also note the importance of human versus machine security. It is arguable that, in many cases, the unit of the trust relationship is neither the human user nor the artificial agent, but, in fact, the whole integrated system, human plus machine. According to the testimony of hacker Kevin Mitnick, it is the human element in a security system that is actually easiest to get round. Computer hackers who enjoy the challenge of getting round a technical fix are obviously a problem, but not necessarily the major security issue (Mitnick 2000)

Complexity of domains is obviously a serious difficulty. In general, any movement towards establishing or spreading trust on the web or semantic web will necessitate a shift from transactions involving centralized information systems to distributed domains and organizations. The picture will be further complicated because all these different information sources will be trusted to varying degrees (Grandison and Sloman 2000). For instance, there are a number of perfectly sensible methods using system-based controls to verify the identity of an agent or a process. If such an agent or process wishes to get access to some resource, then a central database or repository storing access control information can be consulted. However, it is unlikely that such schemes, efficient though they are, will scale up to web dimension applications. For example, authorization cannot, in such circumstances, be divided up into authentication and access control. Therefore, a trust management system cannot rely on centralization and instead must be distributed like the system itself. Credentials provided by agents within the system, such as recommendations and delegations, become essential for making judgements about whether to trust (Kagal et al. 2002).

This is an example of a trust management issue. The question, in such a distributed system, is how to show that an agent is entitled to access to a resource, given the necessary restriction of the information available to the system for verification of the certification (or whatever) that the agent carries with it. Credentials might include agent's properties, for instance, membership of organizations, and delegations from other agents (see Section 5.4).

How might a delegation-based system work? An agent would have the ability to make a delegation of any rights that it possesses to use resources. It also could have the right to delegate delegated to it by another agent. In this case, assuming that the original delegation is legal, the agent could make any delegation that falls within the constraints of the rights it had been granted. A valid delegation will change the access rights of the agents in the system.

Problems might occur in such chains, for example, when agents are created dynamically. An agent may be delegated a right by another agent, which then goes down. When the first agent asks to use the resource and presents its certification, there may be problems if the rights to delegate of

the now non-existent delegating agent cannot be checked. Thus, there may be some requirement for a register of delegation certificates that will either have to be centralized or, alternatively, will need to be stored in some distributed fashion that still allows the owners of resources to get at the certificates easily (and still allows security of the certification process) (Kagal et al. 2002).

Given such a palette of information, trust management issues will include: how to create security policies; what credentials to associate with what properties; and how to reason about credentials and properties to produce a set of decision and access rights. Because such reasoning is a basis for trust management, there will be requirements for such expressive formalisms as trust and security ontologies (Kagal et al. 2002).

Another important trust management issue is that of determining the nature of trust dynamically over time. Current solutions tend not to handle changes in trust, but typically one would expect one's attributions of trust to be altered as knowledge of the world changes, as new actions (which may or may not demonstrate the trustworthiness of agents) occur and as certification and other systems are perceived in different ways. Trust systems need to respond to environmental changes and to learn from agents' behaviour to change trust attributions where this is appropriate. Most current trust applications tend not to incorporate the experience of agents into decision-making (Grandison and Sloman 2000). The value of particular trust policies also needs to be regularly assessed.

6.2 Metadata for Trust

These trust management issues seem to suggest that there is an opportunity for toolkits and languages to aid the expression of and inferences about trust relationships, allowing such relationships to be established, analyzed, evaluated, monitored and reasoned over (Grandison and Sloman 2000). This leads on to a further requisite for trust management, the languages and understanding required to create trustworthiness metadata.

Much of the functionality of the semantic web is premised on the creation of rich sets of metadata about objects – knowledge objects in particular, which will allow machines to reason about their content as much as their external characteristics. Metadata about documents or agents will, by hypothesis, be around in large quantity if the semantic web takes off. In that case, it would be reasonable to expect much of that metadata to be concerned with the trustworthiness of those objects, to allow reasoning about whether a source should be trusted, to what extent, with which resources, for which task.

Such metadata, for example, are likely to be stored in any system recording provenance (see Szomszor and Moreau 2003). The annotations to data that enable the provenance of those data to be inferred are one example of metadata that may be extremely important to collect and curate.

It is arguable that relatively little is known about why certain agents should be trusted, how to recognize trustworthiness and what information is factored in. Much of it will no doubt be similar to the information we have considered in this paper, such as certification of identity or provenance. But equally, there may be other relatively complex factors that principals use in their trust management, and these should be representable via metadata markups. Ontologies of relevant concepts may be available to aid the expression of such metadata and such ontologies are beginning to be seen (for example, Kagal et al. 2002).

We have not considered risk in detail in this paper. The relation between risk and trust, despite a superficial inverse pattern, is not trivial to describe. It seems clear that when calculating which agents or knowledge to trust, and how to mark up a source, risk factors will loom large. These factors include the value of the knowledge under examination, or the value of the services provided by an agent. Value will determine the extent to which trust will be extended; a principal might be more prepared to go out on a limb if the knowledge it was hoping to acquire were more valuable. In particular, value would be closely related to opportunity costs – in other words, what would the principal lose by eschewing the knowledge or services, and how cheap would the knowledge or services be when acquired from another (more trustworthy) source? Is the knowledge or service being obtained by exchange? If so, what is being exchanged and (if it is an excludable good) can the principal afford to lose it? And, will the agent acquire access to any other resources of the principal?

Other factors influencing risk will include more obvious information such as: Who developed the knowledge/provides the service? Who certifies the knowledge? Does the knowledge/service have a wide user base? Is the principal acquainted with anyone in this user base?

It is important that principals should have (i) sufficiently expressive formalisms to enable them to specify precisely what their assessments of their sources are, and (ii) sufficient understanding of their own trust requirements to enable them to make accurate assessments and to interpret metadata provided by other principals (which will probably have different trust requirements). Both of these prerequisites imply at the least the development of useful and used ontologies for expressing trust and for expressing the characteristics of sources which go into a calculation of trust.

6.3 Trust and Trustworthiness

In considering strategies for trust an important problem concerns the creation of a formula for a type of behaviour or for a type of credentials management, in that if it becomes too formulaic there is a danger that it will be too imitable. The issue is the important distinction between trust and trustworthiness (O'Hara 2004a). As Corritore et al. (2003) argue, trust is an act by a principal; trustworthiness is a property of an agent. The agent is in control of its trustworthiness, but is not in control of whether it is trusted; that decision is out of its hands. This gap is essential to the system's working at all. However, if the agent comes to recognize how to influence the decisions of principals to trust it, then it may become able to close, or partially close, that gap.

Put another way, if an agent knows how to signal its trustworthiness, this then takes away all its incentives to be trustworthy. If the metadata assigned to objects, for example, become too fixed or predictable, then the agent may be able to mimic these metadata, to produce them via spurious routes.

Trust management needs to be active and dynamic. It needs managers to be constantly vigilant with respect to potential ways around strategies. Certification authorities and procedures need to be continually updated. Identities need to be monitored. Typical profiles of trustworthy agents or entities need to be treated with care. Equally, if particular strategies for trust management become entrenched (for example, certification of provenance), then such strategies will become overheads on knowledge technologies and web services. They will impose a cost and, in the event that they cease to ensure trustworthiness, that cost will be wasted.

For example, if a certification authority is acting as the means to ensure the provenance of information being used as input by a knowledge technology, then it is important to realise that this imposes *three* overheads on that knowledge technology. First, the knowledge technology must go through a process of verifying the certificate accompanying the information. Second, there will inevitably be trustworthy yet uncertified information (or maybe information certified, but not by a currently recognized authority), which the knowledge technology is unable to use. It is worth remembering in this context that in the fourth quarter of 2002, American online retailers lost an estimated US\$160m through fraud. But they lost an estimated US\$315m through mistakenly rejecting legitimate sales that failed to satisfy rigorous security procedures (Gaudin 2002).

The third overhead is the constant effort that the managers of the knowledge technology must invest to ensure to their own satisfaction that the certification authority was efficient and accurate. As Schneier (2003) argues, overheads such as these must be balanced against a realistic assessment of the costs of failures of trust. There is a danger that trust and security systems become 'motherhood and apple pie' issues while merely imposing needless costs on a system.

6.4 Other Properties of Knowledge Technologies

Similarly, trust has to be balanced against other system properties. The semantic web is at a very early stage of development and the creation of usable techniques for the efficient production of interesting information may be of greater importance than the development of trust mechanisms. At a later stage, when more users need to be brought in to secure large-scale growth for the semantic web, such imperatives may be reversed. Corritore et al. (2003) also argue, as we have noted, that such usability itself contributes to trust. On the other hand, Fogg and Tseng (1999) argue that trustworthiness is a key component of credibility, but it is harder to see how this causative

directionality can be sustained. In addition, the production of useful data abstracted or inferred from large-scale repositories of scruffy data may be inherently hard to achieve in reliable, trustworthy ways.

The development of trustworthy methods of mining data from large repositories is important, but the problem with scaling these methods up to the web scale may not be one of expanding the techniques to deal with more data. The problem may be retaining trustworthiness as the scale begins to outgrow the largest well-managed repositories. Beyond a certain level, data mining and other knowledge extraction techniques might have to work on scruffy and unmaintained datasets, in which case trust may be unable to depend on certificates or evaluations of formal models of reasoning. It may be that trust comes along with usable results, extracted from a wider range of individually less trustworthy datasets.

Using knowledge management techniques and knowledge technologies, however, is a clear way forward to retaining trust while dealing with ever-scruffier datasets.

6.5 Privacy: The Power of Knowledge

Privacy is extremely important for trust. Data will not be released to technologies if those technologies cannot ensure fair use. Here, distinct legal discourses make a difference. A European use-based discourse is relatively restrictive, whereas a rights-based discourse may be more liberal. In the former, the individual can give information away for use for some particular purpose and then the recipient of the information will only be allowed to use it for that purpose. On a rights-based model, owners sign away their rights to control data about them, thereby opening up the possibility of future commoditization of those rights.

The major lines of privacy invasion in the US are based on four specific torts (Camp 2000): i) intrusion upon seclusion; ii) appropriation of name and likeness; iii) false light, which is the publication of misleading information. The information may be true, yet is edited or presented placing the individual in a false light. A link to a commercial web site, written as 'link to a fraudulent site' would, assuming the site to be above board, be an example of this; and iv) public disclosure of private facts.

One serious challenge for privacy advocates with the semantic web and knowledge technologies is the effect of aggregating knowledge. In practical terms, until recently invasions of privacy were monopolized by government and the media, because collection and publication of information were tough barriers to entry. New technologies have changed this radically. Surveillance is cheap and publication is trivial.

Data compilation is much more powerful than merely extracting information from individual datasets. Information about people's incomes may be harmless; information about people's tax records may be harmless. But put the two together and much criminal activity might be exposed.

The US has developed a Code of Fair Information Practice (Camp 2000), which sets out what is reasonable for the management of compilations and data collections. Examples of its recommendations are:

- Data compilations should not be secret;
- Individuals should have access to data collected about them;
- Individuals should be able to audit and correct data; and
- Individuals should be able to prevent disclosure. Prevention of disclosure should be the responsibility of the organization with possession of the data.

Strangely, this code does not as yet apply to medical data. In the face of increasing pressure for data collection and the greater possibilities of data gathering and storage (New Scientist 2003) as studied, for example, in the Memories For Life Grand Challenge submission,⁹ policies and technologies for ensuring privacy are becoming essential, and will be a key research area in the next few years.

7 OPPORTUNITIES FOR INTERDISCIPLINARITY

Trust is first and foremost a social phenomenon, and sociological studies are essential, first to map the properties and functions of trust (Misztal 1996) and second to detail the relationships between people and technology. It should be noted, however, that sociological monographs often have more or less overt political agendas behind them. Many discussions of trust extol the virtues of democracy, consultation, ownership and transparency, all of which are important, but all of which equally are ideologically enshrined in anti-authoritarian and anti-elitist thought, which is aimed explicitly at 'empowerment of the individual' and discounts the possibility that such empowerment might actually be detrimental to some individuals on some occasions (O'Hara 2004a). On the other hand, those discussions of trust that focus on social capital (Fukuyama 1995; Putnam 2000) are often driven by a neo-conservative agenda that is intended to minimize the participation of government in social life, arguing that formal governmental direction of informal social relations will of necessity undermine those very relations (O'Hara 2004a).

Another disciplines that focuses on the relationship between people and technology is management science, including organizational science, behavioural science and other flavours. Ethnography also is important for mapping behavioural traits and is being used increasingly in a number of knowledge technology and semantic web-relevant projects (Cheverst et al. 2001). Psychology also joins the group of related disciplines that mine this seam. Semantic web technologies, such as ontologies and tools, need to be integrated into working practices in order to be taken up (Buckingham Shum 2004; Ellman 2004). Much trust will come 'for free' along with successful integration.

Philosophical analysis is interesting, particularly in the development of ideal theories of trust, which may feed into the calculi of trust that we have discussed. Furthermore, as trust is increasingly seen as an epistemological problem and/or strategy (Luhmann 1979), branches of philosophy such as epistemology, or scientific methodology, will make important contributions – though epistemology will need to be de-psychologized and made relevant to organizational contexts and artificial agents (O'Hara 2002a).

Linguistics has already proved extremely important in the development of new semantic web tools. Being able to extract usable information (knowledge) from the large legacy of plain texts that is available on the web is already proving to be of crucial importance in enhancing the value of that knowledge. Interpretation of such texts is often contested and, therefore, constitutes a dangerous moment for trust; such natural language technologies may improve matters. We have also argued that interrogation and dialogue, using technology to give the written medium of the web some of the properties of the spoken, will aid the development of trust. Linguistics has provided many interesting analyses of the effects of technology on language modes, and these are of major importance to this topic (Ong 1982).

Trust is based to a large extent on incentives, and economics provides perhaps the most intensive study of incentives. Economics, and related fields such as rational choice theory and game theory, should be able to provide theories of how to signal to individuals that trust is in their interests and how to devise systems in which trust is in the interests of the actors (see review by Cave at www.foresight.gov).

Also, needless to say, computer science, cryptography and mathematics have a large part to play by providing the infrastructure for secure systems. Though, as we have argued, technological approaches finesse the problem of trust without necessarily solving it; users must still trust the technological fixes and the motives of those who pay for and police the new systems. In the terminology of O'Hara (2004a), global trust requires underlying local trust.

The reader will no doubt have spotted a number of areas where other disciplines have a role to play. In the following, we highlight areas where such opportunities lie. It should be noted that the scope of many of these extends beyond the area of knowledge technologies.

The following sets out some of the important general opportunities for interdisciplinarity.

7.1 Online and Offline Trust

We have made the assumption, common to many (Corritore et al. 2003), that online trust was

broadly analogous to offline trust. This is an inviting idea, and there is no doubt that many of the properties of the two are the same. However, even ignoring the obvious point that offline trust is a highly heterogeneous social phenomenon (O'Hara 2004a), there are reasons for subjecting the assumption to a deeper examination. This is a project that would require at the least a deep conversation with those whose work it is to study offline trust, for instance, anthropologists, sociologists, economists and philosophers.

In the first place, the Internet as a space has a number of properties broadly analogous to those of the 'real' world, but it also has properties that are not matched by the real world, such as a dramatic fluidity of identity of its denizens. To an extent, the nature of a space will determine the interactions that go on within it and, therefore, the differences between the Internet and other spaces may affect the trust relations that take place within it, making them importantly disanalogous to those of offline trust. An interesting line of research would be to map out the analogies and disanalogies of online and offline trust, and then to determine the significance of the disanalogies.

Secondly, in offline trust the agents involved are of course human. Online trust may involve: two human agents; a human and an artificial agent; two artificial agents; or one or more artificial proxies for human agents.

Clearly, the artificial element will have an effect on how analogous online and offline trust are. But the interesting options opened by the externalization of psychological faculties should not be discounted – the Internet appears to be having a similar effect to that of the spread of literacy and, because artificial agents become involved, that does not necessarily mean that 'trust' is an inappropriate concept here.

In short, the use of the term 'trust' for these online relations is obviously metaphorical. Over-extending a metaphor can be disastrous if we take it too seriously. But equally, metaphors do suggest interesting correspondences that often can be detected in the metaphorical situation. The use of the term 'trust' in an online context has been fruitful so far and we should beware of being too cautious. This is a clear case of where interaction with sociology and related disciplines could prove to be very interesting.

Thirdly, as well as comparing and contrasting online and offline trust, we should also realize that the combination of online and offline trust into a single integrated technological system is also worthy of study. Sociologists of technology are interested in these issues. Important issues include the interface between the online and the offline aspects of the total system, how the online trust serves the purposes of the offline actors, how the assumptions made offline affect the online interactions, and how either the online or the offline trust mechanisms could override one another in any integrated application.

7.2 Trust as Utilitarian versus Trust as a Mark of a Moral Community

In the sociology literature, two related but different concepts of trust seem to be in operation. The first is a utilitarian concept, associated particularly with Niklas Luhmann (1979). This type of trust is an acceptance of collaborators' *bona fides*. Thus, the principal makes tangible gains from trusting. It cuts transaction costs, as it no longer has to investigate the agent or perform the subtask itself. It is therefore rational for the principal to trust. The direction of causality is from the agent to the principal; the agent's good behaviour leads the principal to trust it. The good behaviour is the cause of the trust; the trust is the effect of the good behaviour. This form of trust is most commonly associated with economic domains, such as e-commerce, and has been explicitly associated with online trust (Grabner-Kräuter and Kaluscha 2003).

The second is a more value-based trust concept, which comes about through a consensus of values. This is a more traditional notion of trust going back to Durkheim (1893) and Parsons (1949), although empirical analyses of survey data have been argued to support the assumption that this notion of trust is still prevalent in society (Uslaner 2002). Here, to trust someone is to accept them into one's moral community, where there may be fewer rational justifications for this to happen – indeed, to do so may entail quite some risk for the principal. The direction of causality is precisely reversed – the principal's trust leads an agent to good behaviour. The trust is the cause of the good behaviour, the good behaviour the effect of the trust.

In the context of the Internet, this second notion of trust is interesting because – almost uniquely among technologies – the Internet is very value-laden. Hackers have a set of values that lead them to attempt to disrupt normal operations; they do not do it for gain in most cases (O'Hara 2004a). Scientists and academics have a knowledge-sharing, public-spirited set of values that have fostered, for example, the e-prints movement (Harnad 2003). On the other hand, Microsoft pioneered the idea of selling software by restricting access to source code, protecting and therefore creating the option of exploiting their intellectual property. The open source movement claims to produce better software, while exposing source code to view.

All these decisions are based on a set of values, and it is not hard to characterize Internet politics as a series of clashes of these value sets (see Lessig 1999). It would be interesting, therefore, to investigate, together with sociologists and other social scientists, which, in this context, would be the best notion of trust to deal with.

In addition, there are several opportunities for interdisciplinary research that are specific to the knowledge technologies field.

7.3 The Nature of Knowledge Acquisition

The clear connection for knowledge technologies is with the field of epistemology. There are three major types of interaction that would be useful as a focus for research.

First, trustworthy knowledge technologies require trustworthy processes for producing knowledge from the input information. Therefore, the field of analytic epistemology should provide a useful set of guidelines for when knowledge acquisition processes are trustworthy – and collaboration could be envisaged with philosophers of science and theorists of probability.

Second, knowledge is also very powerful; possession of it (particularly monopolistic possession), or the perception of possession, can lead to the owner having a great deal of power. A continental tradition of epistemology, based, for example, on the writings of such as Habermas ([1968] 1987) and Foucault and Lyotard (Lyotard [1979] 1984), while often anti-technology, provides important commentaries on the way knowledge is used to get things done and the legitimacy of the power thus created. These issues are very relevant to trust.

Third, applied epistemology, or the use of technology to create knowledge in a systematic way, has been a feature of the software industry for some time. For example, the field of knowledge acquisition was associated with the creation of knowledge bases for expert systems and the like from the late 1980s onwards. Knowledge acquisition, data mining and machine learning can shine very interesting light on the possibilities for knowledge technologies. At a higher level of abstraction, the areas of cognitive psychology that crucially informed these technologies are also of interest.

7.4 The Nature of Personalized Interactions

Interrogation and dialogue are important for establishing trust. We did not speculate in Section 5.8 on why this might be – is it the responsiveness of speech to interaction, or is it the greater bandwidth of communication that presence allows? However, if it would be possible, via collaboration with linguists or anthropologists, to establish the key trust-creating or trust-enhancing properties of oral communication over literate communication, then this would be of great help in designing semantic web tools and knowledge technologies that exploit those key properties.

7.5 Brands, Reputations and other Filtered Narratives

Selection of interlocutors is important, and if knowledge technologies can take input only from trustworthy sources, then we will be a long way towards ensuring the production of trusted knowledge from scruffy web scale sources. The main way to do this is to create a reputation, that is, a filtered narrative taken from past events involving the source, that signals trustworthiness (or untrustworthiness). Reputations can be fair or unfair, but they are important in reducing transaction costs (O'Hara 2004a).

Particular symbols of reputation that are very important repositories of trust are *brands* (O'Hara

2004a), and restricting knowledge acquisition to branded websites is obviously an important move, although it may be too restrictive in some circumstances, where the relative quantity of branded websites is low. Nevertheless, an interesting research collaboration would be with those disciplines studying branding as a phenomenon and as a mechanism for spreading trust, for example, economics and business studies.

7.6 Effective Maintenance of Knowledge Bases

Finally, effective maintenance of knowledge bases and information sources helps of course to retain their trustworthiness. Knowledge maintenance is notoriously difficult because it involves a substantial overhead. Yet if maintenance became more of an accepted practice, a cost that information-providers were more generally prepared to take on, then, in general, the amount of trustworthy knowledge available on the Internet would increase, and large-scale knowledge harvesting efforts would be that much more reliable.

An obvious set of collaborations, therefore, would be with the management science and knowledge management communities to try to establish sensible methodologies for knowledge maintenance that could be integrated with standard work practices without imposing too much of an overhead.

8 CONCLUSION

We have discussed a number of issues and open questions with respect to trust, its creation, propagation and conservation on the Internet and the World Wide Web, particularly bearing in mind likely technological developments to extend the web. We have set out our understanding of trust and the likely technological developments of the semantic web and knowledge technologies.

Trust can be viewed at a micro or macro level. At the micro level, a series of tactics can, in various circumstances, help create or preserve trust. At the macro level, such tactics need to be combined into trust strategies. Various tactics were set out, some of which are variants on others. For example, there are many variations on the tactic of restricting those sources of knowledge that a knowledge technology uses, including relying on branded websites, and demanding verifiable certification of provenance. Managing trust is a key managerial requirement for the semantic web, and an interesting demand that has come to light is for informative metadata about knowledge sources that can be used for assessing trustworthiness.

We have also discussed the contribution that can be made by various disciplines to the issues raised in this paper. We highlighted a number of interesting potential research directions and ways of developing, monitoring and maintaining trust online. Several research strands look particularly promising.

Metadata and provenance. The creation and curation of the metadata that signal trustworthiness and provide evidence of a piece of information's provenance are very important. Issues to be addressed include the identification (possibly dynamically) of the metadata most relevant to trust issues in a particular context and methods of storing these metadata in such a way as to maximize both security and ease of access for queries.

Trust management. Here the issues include how to provide a system that can deal with the trust issues as the agent communities scale up. Is it possible, for example, to keep track of a series of agents' delegation rights without a centralized authority that is likely to be overwhelmed by requests in an environment of the scale of the web?

Transitivity. As we noted, trust is not strictly transitive. However, networks of actors making recommendations can be a very powerful method of spreading trust. Key issues include how to understand this process; trust in recommendations will naturally decay as the chains of recommendations shrink, but the rate of decay may well change from application to application.

Interrogation. Trust is enhanced when an agent can be interrogated, through dialogue, as opposed to merely presenting certificates that provide a fixed set of credentials. For knowledge technologies on the semantic web, such tactics are particularly plausible.

Knowledge management. Finally, given that we are focusing on knowledge technologies, managing knowledge effectively and presenting it in a timely way will have a strong effect on trust, even if that effect is only indirect. Various data management housekeeping issues will enhance

both trust (because a principal can see that knowledge is being curated accurately and processed efficiently) and trustworthiness (because the technology's information processing should be more reliable).

APPENDIX 1 THE SEMANTIC WEB¹⁰

The World Wide Web is based on HTML (the Hypertext Markup Language). Web pages are created by commands in HTML. Because HTML is the common web language, web pages are always more or less alike, and can all be seen by the main browsers. HTML tells the computer how to arrange all the information in a web page on a screen, where the text should go, how it should be formatted, where the pictures fit, how the different panels of text should fit together, and so on; it controls the look of the page.

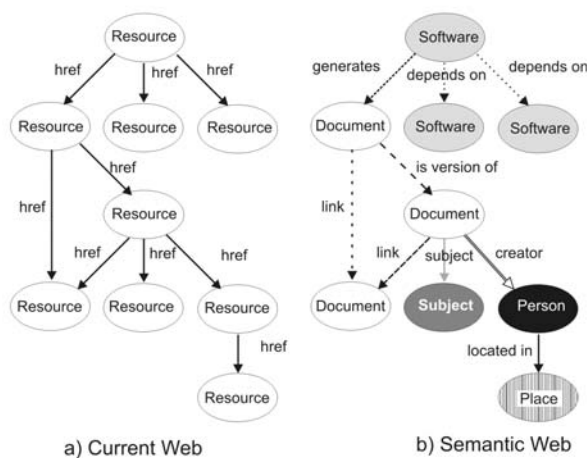
Navigation through the web is controlled by hyperlinks and search engines; however, this system is reaching the limits of its capacity, given that there are more than half a trillion web documents. The more information that appears on the web, the less use a simple key word system, as used by search engines, is; it will give you too many irrelevant hits. The trouble with key words is that they are *uninterpreted*; if your key word is, say, Bush, a key word system is unable to tell the difference between George Bush, Kate Bush, the African Bush or the metal lining of an axle-hole (known as a bush, believe it or not). Therefore, the semantic web is being developed.

Playing the HTML role in the semantic web is a language called XML, the eXtensible Markup Language. XML differs from HTML in that it allows users to define little sub-languages for describing objects. Whereas HTML tells the computer how to arrange the content on the page, XML allows you to tell the computer what the things named by the content are.

However, this is not enough. The computer can do very little more with that information than it could do before. Therefore the semantic web needs the Resource Description Framework (RDF). RDF is a framework that brings together three things, two objects, and a relation between them. So, for example, the two objects might be 'University of Southampton', and a picture of the university buildings (which will be a file, perhaps called something like 'soton.jpg'). RDF enables you to tell the computer that there is a relationship between the two: the relation might be called 'picture-of', and so RDF lets us assert that 'soton.jpg picture-of "University of Southampton"' – that is, that the picture is of the University of Southampton.

XML lets you tell your computer what things are, and RDF lets you tell it how these things are related. A third element is still needed, because the computer still does not know (in a metaphorical sense of 'know') what these terms mean. This is achieved with the use of *ontologies*.

Figure A 1-1 From the World Wide Web to the semantic web



Source: Adapted from Koivunen and Miller (2002, p. 30).

An ontology is a specification of the language and concepts of a restricted domain of discourse, and gives as it were the conceptual background to the words and phrases used in XML and RDF. These words and phrases are therefore defined in terms of each other, which specifies a little web of interrelated concepts. The technology is too weak to say that the computer understands anything (in other than metaphorical senses), but an ontology will provide the beginning of an interrelated web of terms in the sense of Quine and Ullian (1970). The computer could be told, for instance, that a university is a type of *educational establishment*, that it has *students* and *lecturers*, an *address*, a *website*, a *telephone number*, that the lecturers will include *professors*, *senior lecturers* and *readers*, that the students will include *undergraduates* and *postgraduates*, that the postgraduates will have *degrees*, that these degrees will be in *subjects*, and will have been awarded by *educational establishments*, and so on. The ontology links together all the concepts and terms that will help the computer to make holistic sense of the terms like 'university' that were being used in the XML specification of the webpage.

In short, whereas in the World Wide Web, HTML told the computer how to arrange the content on the screen, in the semantic web, XML + RDF + ontologies tell it not only how to arrange the content, but also *what it is all about*. The effect is suddenly to provide the computer with a richer characterization of the domain. Figure A 1-1 shows how the new expressivity can allow the computer to see much more.

On the left, we see the basic web view of the computing world; the resources are whatever is held at web addresses, and they are connected by hypertext links written in HTML. But on the right, the same domain is seen in much more detail. With an XML characterization of the domain, the computer can see that some of the resources are pieces of software, others are documents, others are persons and so on. The links are made more meaningful too by RDF. For instance, we can see that one document is a version of another, that the creator of one of the documents is a particular person, and so on. The formalisms of the semantic web allow you to tell the computer so much more about the domains you are describing.

If we now look at what is becoming the increasingly well-known layered model of the semantic web (Figure 3), how in the light of this discussion do we interpret these layers? At the bottom are Unicode (a standardised system for encoding data) and URIs (Uniform Resource Identifiers – addresses of resources). These are the nuts and bolts of the semantic web. But there is no point having these unless you have XML to tell you what they refer to. And there is no point having XML unless you have RDF to tell you how those things relate to each other. Similarly, there is no point having XML and RDF unless you have ontologies to explain the significance of the XML classes and RDF relations. There is no point in having ontologies without a logic to provide methods of inferring one thing from another. There is no point inferring things without a theory of proof to tell you that the inferences are sound.

And, finally, there is no point in having a system of proof unless those who will use it (in the case of the semantic web, this means the 654 million Internet users) have confidence in it. Without the trust of the users in the system, the semantic web will never get off the ground.

NOTES

- 1 Thanks are due to anonymous referees for several improvements to this paper.
- 2 See Misztal (1996) for a review of some of the best-known approaches.
- 3 See also the large empirical literature that they cite.
- 4 See www.aktors.org, accessed 17 Apr. 04.
- 5 See <http://rdfweb.org/foaf/>, accessed 17 Apr. 04.
- 6 See <http://www.isi.edu/ikcap/trellis/>, accessed 17 Apr. 04.
- 7 There are some exceptions to this, such as PolicyMaker, which binds access rights to the public key, see Blaze et al. (1996).
- 8 A similarly-inspired algorithm is exploited by Richardson et al. (2003).
- 9 See <http://www.csd.abdn.ac.uk/~ereiter/memories.html>, accessed 17 Apr. 04.
- 10 Adapted from O'Hara (2004b).

REFERENCES

- Alani, H., Dasmahapatra, S., Gibbins, N., Glaser, H., Harris, S., Kalfoglou, Y., O'Hara, K. and Shadbolt, N. (2002), 'Managing Reference: Ensuring Referential Integrity of Ontologies for the Semantic Web' in A. Gómez-Pérez and V.R. Benjamins (eds) *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, Berlin: Springer-Verlag, pp. 317-34.
- Alani, H., Dasmahapatra, S., O'Hara, K. and Shadbolt, N. (2003a), 'Identifying Communities of Practice through Ontology Network Analysis' *IEEE Intelligent Systems* Mar/Apr: 18-25.
- Alani, H., Kim, S., Millard, D., Weal, M., Hall, W., Lewis, P. and Shadbolt, N. (2003b), 'Automatic Ontology-based Knowledge Extraction from Web Documents', *IEEE Intelligent Systems* (18(1)): 14-21.
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001), 'The Semantic Web' *Scientific American* May: 34-43.
- Blaze, M., Feigenbaum, J. and Lacy, J. (1996), 'Decentralized Trust Management' in *Proceedings of the 1996 IEEE Symposium on Security and Privacy*, Washington DC: IEEE Computer Society, pp. 164-73.
- Buckingham Shum, S. (2004), 'Contentious, Dynamic, Information-sparse Domains. And Ontologies?' *IEEE Intelligent Systems* Jan/Feb: 80-81.
- Buckingham Shum, S. and Sumner, T. (2001), 'JIME: An Interactive Journal for Interactive Media', *Learned Publishing* 14(4): 273-85.
- Camp, L.J. (2000), *Trust and Risk in Internet Commerce*, Cambridge MA: MIT Press.
- Carr, L., Bechhofer, S., Goble, C. and Hall, W. (2001), 'Conceptual Linking: Ontology-based Open Hypermedia', in *Proceedings of the 10th World Wide Web Conference*, New York: ACM Press, pp. 334-42.
- Cheverst, K., Clarke, K., Cobb, S., Hemmings, T., Kember, S., Mitchell, K., Phillips, P., Proctor, R., Rodden, T. and Rouncefield, M. (2001), 'Design with Care' *New Technology in Human Services* 14(1/2): 39-47.
- Corritore, C.L., Kracher, B. and Wiedenbeck, S. (2003), 'On-line Trust: Concepts, Evolving Themes, A Model', *International Journal of Human-Computer Studies* 58(6): 737-58.
- Dimitrakos, T. and Bicarregui, J. (2001), 'Towards Modelling e-Trust' presented at the 3rd Panhellenic Logic Symposium, Anogia, Greece,
http://www2.nr.no/coras/workshop_at_RAL/Dimitrakos_Modelling_Trust/sld022.htm accessed 17 Apr. 2004.
- Domingue, J., Motta, E., Buckingham Shum, S., Vargas-Vera, M. and Kalfoglou, Y. (2001), 'Supporting Ontology-driven Document Enrichment within Communities of Practice', in *Proceedings of the International Conference on Knowledge Capture*, New York: ACM Press, pp. 30-37.
- Durkheim, E. (1893/1984) *The Division of Labour in Society*, London: Palgrave Macmillan.
- Ellman, J. (2004), 'Corporate Ontologies as Information Interfaces', *IEEE Intelligent Systems* Jan/Feb: 79-80.
- Fogg, B.J. and Tseng, H. (1999), 'The Elements of Computer Credibility' *Proceedings of CHI 99*,
http://www-pcd.stanford.edu/captology/Key_Concepts/Papers/Credibility25.PDF, , accessed 17 Apr. 04.
- Fukuyama, F. (1995), *Trust: The Social Virtues and the Creation of Prosperity*, New York: Free Press.
- Gaudin, S. (2002), 'Online Fraud Growing in Scale, Sophistication' *Datamation*, 5 Dec,
<http://itmanagement.earthweb.com/ecom/article.php/1552921>, accessed 17 Apr. 04.

- Gil, Y. and Ratnakar V. (2002), 'Trusting Information Sources One Citizen at a Time' in I. Horrocks and J.A. Hendler (eds) *Proceedings of the First International Semantic Web Conference*, Berlin: Springer, pp. 162-76.
- Golbeck, J., Hendler, J. and Parsia, B. (2003), 'Trust Networks on the Semantic Web' in M. Klusch, S. Ossowski, A. Omicini and H. Laamanen (eds) *Proceedings of Cooperative Information Agents VII (CIA 2003)*, Berlin: Springer, pp. 238-49.
- Grabner-Kräuter, S. and Kaluscha, E. (2003), 'Empirical Research in On-line Trust: A Review and Critical Assessment', *International Journal of Human-Computer Studies* 58(6): 783-812.
- Grandison, T. and Sloman, M. (2000), 'A Survey of Trust in Internet Applications' *IEEE Communications Surveys* 4th Quarter,
www.comsoc.org/livepubs/surveys/public/2000/dec/grandison.html, accessed 17 Apr. 04.
- Habermas, J. ([1968] 1987), *Knowledge and Human Interests*, (translated by Jeremy J. Shapiro), Cambridge: Polity Press.
- Hardin, R. (1999), 'Do We Want Trust in Government?', in M.E. Warren (ed.) *Democracy and Trust*, Cambridge: Cambridge University Press, pp. 22-41.
- Harnad, S. (2003), 'Eprints: Electronic Preprints and Postprints' in M. Dekker (ed.) *Encyclopedia of Library and Information Science*,
<http://eprints.ecs.soton.ac.uk/archive/00007721/01/eprints.htm>, accessed 17 Apr. 04.
- Hu, Y.-J. (2001), 'Some Thoughts on Agent Trust and Delegation', in *Proceedings of Autonomous Agents 2001*, New York: ACM Press, pp. 489-96.
- Kagal, L., Finin, T. and Joshi, A. (2002), 'Developing Secure Agent Systems using Delegation Based Trust Management' in K. Fischer and D. Hutter (eds) *Proceedings of 2nd International Workshop on Security in Mobile Multiagent Systems (SEMAS 02), held with Autonomous Agent and Multiagent Systems (AAMAS-2002)*, Research Report RR-02-03, Kaiserslautern: Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, pp. 27-34.
- Koivunen, M.-R. and Miller, E. (2002), 'W3C Semantic Web Activity', in E. Hyvonen (ed.) *Semantic Web Kick-Off in Finland – Vision, Technologies, Research and Applications*, Helsinki: Helsinki Institute for Information Technology, pp. 27-44.
- Lessig, L. (1999), *Code and Other Laws of Cyberspace*, New York: Basic Books.
- Li, Y. and Mascagni, M. (2003), 'Improving Performance via Computational Replication on a Large-scale Computational Grid', in *Proceedings of the 3rd IEEE International Symposium on Cluster Computing and the Grid (CCGRID2003)*, New York: IEEE Computer Society, pp. 442-49.
- Luhmann, N. (1979), *Trust and Power*, Chichester: John Wiley & Sons.
- Lyotard, J.-F. ([1979] 1984), *The Postmodern Condition: A Report on Knowledge*, (translated by G. Bennington and B. Massumi), Manchester: Manchester University Press.
- Malhotra, D. and Murnighan, J.K. (2002), 'The Effects of Contracts on Interpersonal Trust' *Administrative Science Quarterly* Sept.: 534-59.
- McDermott, R. (1999), 'Why Information Technology Inspired but Cannot Deliver Knowledge Management' *California Management Review* 41: 103-17.
- Misztal, B.A. (1996), *Trust in Modern Societies*, Cambridge: Polity Press.
- Mitnick, K. (2000), Testimony to US Senate Committee on Governmental Affairs, March,
http://www.senate.gov/~gov_affairs/030200_mitnick.htm, accessed 17 Apr. 04.
- New Scientist (2003), 'The Story of your Life ... On a Laptop', *New Scientist* 4 Oct., p. 28.
- O'Hara, K. (2002a), *Plato and the Internet*, Cambridge: Icon Books.
- O'Hara, K. (2002b), 'The Internet: A Tool for Democratic Pluralism?' *Science as Culture* 11(2): 287-98.
- O'Hara, K. (2004a), *Trust: From Socrates to Spin*, Cambridge: Icon Books.

- O'Hara, K. (2004b), 'Trust, Socrates and the Internet' in M. Toyota and J. Noguchi (eds) *Speech, Writing and Context: Interdisciplinary Perspectives*, Osaka: Intercultural Research Institute of Kansai Gaidai University (in press)
- Ong, W.J. (1982), *Orality and Literacy: The Technologizing of the Word*, London: Methuen and Co.
- Parsons, T. (1949), *Structure of Social Action*, Glencoe IL: Free Press.
- Povey, D. (1999), *Trust Management*,
<http://security.dstc.edu.au/presentations/trust/>, accessed 17 Apr. 04.
- Putnam, R. (2000), *Bowling Alone: The Collapse and Revival of American Community*, New York: Simon and Schuster.
- Quine, W.V. and Ullian, J.S. (1970), *The Web of Belief*, New York: Random House.
- Richardson, M. Agrawal, R. and Domingos, P. (2003), 'Trust Management for the Semantic Web' in D. Fensel, K.P. Sycara and J. Mylopoulos (eds) *The Semantic Web – ISWC 2003, Second International Semantic Web Conference*, Berlin: Springer, pp. 351-68.
- Schneier, B. (2003), *Beyond Fear: Thinking Sensibly About Security in an Uncertain World*, New York: Copernicus.
- Shadbolt, N., O'Hara, K. and Crow, L. (1999), 'The Experimental Evaluation of Knowledge Acquisition Techniques and Methods: History, Problems and New Directions', *International Journal for Human-Computer Studies* 51(4): 729-55.
- Shadbolt, N. R., Gibbins, N., Glaser, H., Harris, S. and schraefel, m.m.c. (2004 forthcoming), 'CS AKTive Space: Or How We Stopped Worrying and Learned to Love the Semantic Web' *Proceedings of the Second International Semantic Web Conference, IEEE*, New York: IEEE Computer Society.
- Shelat, B. and Egger, F.N. (2002), 'What Makes People Trust Online Gambling Sites?' CHI 2002 Abstracts,
<http://www.ecommuse.com/research/publications/chi2002.pdf>, accessed 17 Apr. 04.
- Spiekermann, S., Grossklags, J. and Berendt, B. (2001), 'E-privacy in 2nd Generation e-Commerce: Privacy Preferences versus Actual Behavior' *Proceedings of the 3rd ACM Conference on Electronic Commerce, Tampa*, New York: ACM Press, pp. 38-47.
- Stanford, J., Tauber, E.R., Fogg, B.J. and Marable, L. (2002), 'Experts vs Online Consumers: A Comparative Credibility Study of Health and Finance Web Sites' *Consumer WebWatch*, 29 Oct.,
http://www.consumerwebwatch.org/news/report3_credibilityresearch/slicedbread_abstract.htm, accessed 17 Apr. 04.
- Sunstein, C. (2001), *Republic.com*, Princeton: Princeton University Press.
- Szomszor, M. and Moreau, L. (2003), 'Recording and Reasoning over Data Provenance in Web and Grid Services' in R. Meersman, Z. Tari and D.C. Schmidt (eds) *On the Move to Meaningful Internet Systems 2003: CoopIS, DOA and ODBASE – OTM Confederated Conferences*, Berlin: Springer, pp. 603-20.
- Tennison, J., O'Hara, K. and Shadbolt, N. (2002), 'APECKS: Using and Evaluating a Tool for Ontology Construction with Internal and External KA Support' *International Journal for Human-Computer Studies* 56(4): 375-422.
- The Economist (2003), 'The New Geography of the IT Industry' *The Economist*, 19 July.
- Uslaner, E.M. (2002), *The Moral Foundations of Trust*, Cambridge: Cambridge University Press.
- Wenger, E. (1998), *Communities of Practice: Learning, Meaning and Identity*, Cambridge: Cambridge University Press.
- Winslett, M., Yu, T., Seamons, K.E., Hess, A., Jacobson, J., Jarvis, R., Smith, B. and Yu, L. (2002), 'Negotiating Trust on the Web', *IEEE Internet Computing* Nov/Dec: 30-37.