3. Untangling Causality in Design Science Theorising

Dirk S. Hovorka
Bond University

Shirley Gregor
The Australian National University

Abstract

Although design science research aims to create new knowledge through design and evaluation of artefacts, the causal agency through which artefacts obtain predicted outcomes is frequently under-specified. Within this domain of knowledge, six types of causal reasoning can be applied by researchers to more clearly articulate why desired outcomes will result from the implementation of the artefact. In addition, reflecting on the causal foundations of the design will enable more definitive evaluation of the design theory and scientific explanation of the behaviour of the artefact-in-use. The framework proposed here is based on an extensive literature in causal theory and the implications are that researchers will be able to articulate the causal reasoning used in design science theorising.

Introduction

Design science research (DSR) seeks to create new knowledge through the process of designing, building and evaluating information system artefacts. Designed systems are teleological in nature: they have an intended purpose, and designers and users have expectations of specific observable outcomes as a direct result of implementation and use. The purpose of design lies in shaping artefacts and events to create a more desirable future (Boland, 2002). As these systems are intended to mediate or intervene in personal, group or organisational activities to produce specific outcomes, they are perceived to have causal agency, either implicitly or explicitly. Any proposed design solution prescribes technological rules that provide general instructions for building an artefact intended to produce a specific outcome (Bunge, 1967; van Aken, 2004).

In addition, these technological rules can form part of a design theory that links specific architectures and outcomes and, additionally, predicts and explains the outcomes that obtain from those structures. The suggestion that DSR is intended to contribute to our theoretical knowledge (Gregor, 2006; Gregor and Jones, 2007; Venable, 2006) has become more generally accepted. Beginning at least with Aristotle, to know the causes of things was fundamental to the explanatory disciplines and is still characteristic of modern science (Bunge, 2008; Salmon, 1998). Thus, reasoning about causality is required by both the designers of artefacts in their construction activities and development of design theory and the researchers who study the behaviour and effects of artefacts-in-use for evaluation purposes, to inform future design, and in building theory about designed artefacts.

Despite the implicit reliance on causal reasoning and its centrality in theory building, the problem of causality in the DSR literature has been little addressed or has been addressed in a relatively simplistic fashion. Only rarely are causal connections explicitly specified in DSR and, when identified, such connections are only very generally described. In many cases where kernel theories are specified, researchers retreat behind the simplification of cause expressed in statements that a specific kernel theory provides justification for the prediction and explanation of the desired outcomes.[1] Direct connections between the causal mechanisms of the kernel theory and how these causal mechanisms will be instantiated in the design to produce the expected outcome are rarely described. Work that studies artefacts-in-use frequently employs statistical methods, where questions of causality are avoided or glossed over. Although, as Venable (2006) points out, it is possible to create a design that successfully produces a better state of affairs in the problem space, but not know how or why it works. But this limitation constrains the contribution to the knowledge base and our ability to apply that knowledge in other domains.

The extensive discussion of causality in the philosophy of science literature precludes anything but a modest review in a chapter such as this. We draw on this literature, however, to provide a basic conceptualisation of causation and to propose a framework that might guide researchers in the process of design theorising and in the evaluation of artefacts and the knowledge discovered as a result of artefact construction and use. The goal of the framework is to sensitise researchers to causal reasoning in design science research. The framework itself is not prescriptive but can be used to identify and refine causal reasoning as

1 A kernel theory, or 'justificatory knowledge' or 'micro-theory', is the explanatory knowledge that links other components of a design theory: design goals, principles, processes and materials. It is the knowledge that explains 'why' a particular design is expected to work and thus involves causal reasoning (Gregor and Jones, 2007). For example, knowledge from cognitive science explains why certain interface design principles are valid. Kernel theory can come from reference disciplines such as cognitive science or from other design theory (for example, theory from aesthetics).

it is applied to research in the interior prescriptive mode (the construction of artefacts) and also in the exterior descriptive mode (use of artefacts in socio-technical systems) (Gregor, 2009). More explicit causal reasoning will enable design science (DS) researchers to better select and apply kernel theories, to evaluate design principles more deeply and to provide stronger knowledge claims when evaluating socio-technical systems. It will also inform descriptive research of artefacts-in-use, which will aid the development of theory that better informs design science research.

Importance of Causal Thinking in Design Science 'Theorising'

Design is inherently based on causal claims or assumptions. All human activities intended to shape or control future events are based on causal inference and the design is a 'specifications of actions to be taken (often in a specified sequence) to achieve the intended consequence' (Argyris, 1996, p. 396). In theorising about design of information systems (IS), the causal agency may be transferred to the technology and/or to the users' interactions with the technology. In some cases, the causal agency may exist at the level of organisational actions in the implementation or control of the technology.

The IS community has engaged in considerable discussion and argument about the nature and relevance of theory in DSR (Gregor and Jones, 2007; Hooker, 2004; Venable, 2006) and we do not seek to enter this debate in this essay. But whether a researcher produces a set of design principles, a design theory or a set of technological rules, the design of a teleological artefact contains warrants about antecedent–consequent relationships that must be grounded in existing knowledge. Goldkuhl (2004) identified empirical, theoretical and internal types of grounding. Of particular importance in this discussion is the explanatory aspect of the theoretical grounding type that is found in the reliance on kernel theories as a basis for design.

Selection of kernel theories is vexing because in most problem domains multiple, sometimes contradictory, theories exist. The design researcher must select, from the possible theory base, kernels that seem relevant to the design problem often without full knowledge of their suitability to the design problem. As an external source of theoretical grounding, examination of the causal claims in the kernel theories potentially provides a stronger grounding for the resultant design theory. In addition, explicit recognition of the causal commitments assumed in the design become clear research questions for the evaluation phase and might lead to improved knowledge contributions from the DSR process.

Causal reasoning will also enable a better contextualisation by identifying how, when and where kernel theories are applicable and what interactions between kernel theories can be expected.

We note that not all design theorising is based on kernel theories and, even when present, kernel theories might serve as creative inspiration rather than a source of logical derivation of design theory (Goldkuhl, 2004). Causal reasoning requires the researcher to evaluate in what context, and for which system, participants and tasks each specific kernel theory was or was not relevant to the design (Hovorka and Germonprez, 2009). In these instances, the increased ability for appropriate evaluative and knowledge construction is a sufficient argument for description of the *assumed* causal mechanisms.

Background

As a research approach, DSR is based on the idea that knowledge of the world can be obtained through a 'problem identification-build-evaluate-theorise' process (Hevner et al., 2004). In the problem identification phase, researchers identify a problem domain in which the intervention of a new artefact into an environment will produce a change in specified outcomes. This phase inherently involves causal thinking—specific design characteristics of the artefact have causal agency to produce outcomes that will solve the identified problem. Causal reasoning also appears during theory development. As noted by Gregor (2006), a theory that explains, predicts or prescribes is offering a causal explanation by identifying what antecedent conditions will result in specific consequents.

The identification of causality is, however, extremely problematic, and among philosophers the very concept of causation has suffered numerous deaths, including strong critiques resulting from quantum theory and logical positivism (Bunge, 2008). Yet reasoning about causation—whether conceived as complete determinism (that is, Laplace's daemon), as a specific relationship between entities (Salmon, 1998) or as regularities in perception leading to cognitive beliefs (Hume, 2004)—is an instinctive tendency of human behaviour (Bunge, 2008). Humans are curious, and the case can be made that their survival depends on determining why events occur and how to intervene to shape their environments in a desirable way. Yet arguably, design of the artefacts that are the object of study for DSR is founded on many different types of determinism and is lumped into a causal language that distorts the real contributions to knowledge. Although the word 'cause' is often omitted in research papers, our ideas of determination of effects based upon designed artefacts bear the stamp

of causality (Bunge, 2008). Untangling and clarifying precisely how design theory and artefacts determine outcomes will benefit both our design theories and the knowledge created by design science research.

Much of the analytic thinking about causation is based on the assumption that individual causes are independent of each other—that changes in one factor will affect the outcome but will not change any other factor. Yet this assumption does not hold in the real-world situations in which DSR operates. Moreover, the model of the world assumed in DSR is one of general linear reality (Abbott, 2001), in which the order of events does not influence the outcomes. Yet the outcomes resulting from the artefacts in use vary with time and context, the order in which information is presented affects human decisions and causal agency changes for different stakeholders over time.

The concept of causality has a long history that can be traced back to Aristotle and the early Greek philosophers, who recognised a fundamental distinction between descriptive knowledge saying that something occurred and explanatory knowledge saying why something occurred. Notably, Aristotle's doctrine identified four causes (*aitia*) (from a translation by Hooker, 1996):

- *material cause*: 'that out of which a thing comes to be, and which persists' (that is, what a thing is made of)
- *formal cause*: 'the statement of essence' (that is, the form and pattern that define something as 'this' rather than 'that')
- *efficient cause*: 'the primary source of change' (that is, the designer or maker of something)
- *final cause*: 'the end (*telos*), that for the sake of which a thing is done' (for example, health is the cause of exercise).

Modern science has focused primarily on efficient causes, the agents that bring about change, with material causes assumed to be that which is changed.

There are numerous means of reasoning about causality and at different times and contexts we might ascribe causality through different theoretical conceptions. Practical criteria for determination of causality were presented by J. S. Mill (1882) as: 1) the cause has to precede the effect in time; 2) the cause and effect must be related; and 3) other explanations of the cause–effect relationship have to be eliminated (Shadish et al., 2002). These criteria are still relevant, but they are overly simplistic when dealing with the construction of information technology-based artefacts.

For the purpose of an analytic framework, we begin with consideration of two important views of causation: *event causation* and a form of *agent causation* (largely following Kim, 1999).

Event causation is the causation of an event by some other event or events. A computer virus or power outage will cause system disruption, but the opposite is not true. Kim (1999) distinguishes four prominent approaches to analysing event causality.

1. *Regularity analysis* (constant conjunction or nomological analysis): This is the type of causality common in the natural sciences and is based on uniform and constant covering laws. 'There are some causes, which are entirely uniform and constant in producing a particular effect; and no instance has ever been found of any failure or irregularity in their operation' (Hume, 2004, p. 206). It is argued, however, that due to the complexity and variability of human behaviour, this type of regularity should not be expected or sought in the social sciences (Fay, 1996; Little, 1999).
2. *Counterfactual analysis:* This means of analysis posits that what qualifies an intervention as a cause is the fact that if the intervention had not occurred, the outcome would not have happened (the cause is a necessary condition). To say that striking a drinking glass caused the glass to break is to say that the breaking was counterfactually dependent on the strike. If the glass had not been struck it would not have broken (*ceteris paribus*) (Collins et al., 2004).
3. *Probabilistic analysis:* This type of causality was recognised by Hume (2004, p. 206)—in comparison with universal laws, 'there are other causes, which have been found more irregular and uncertain; nor has rhubarb always proved a purge, or opium a soporific to everyone, who has taken these medicines'. This view of causal analysis is thought to be suited to the social sciences, where the lack of a closed system and the effects of many extraneous influences make other causal analysis difficult to undertake. 'To say that C is the cause of E is to assert that the occurrence of C, in the context of social processes and mechanisms F, brought about E, or increased the likelihood of E' (Little, 1999, p. 705).
4. *Manipulation analysis:* This conception of causation entails the idea that an intervention in a system will influence the outcomes. That is, the cause is an event (an act) that we can manipulate or perform to bring about an effect—for example, pressing a switch turns a light off. This practically oriented conception can identify knowledge useful for specific kinds of prediction problems. It contains elements of variance such that probabilistic effects can be accounted for. More importantly, it provides a separate inferential step that allows us to differentiate the case where two variables are correlated, from the case where it is claimed that one variable will respond when under manipulation by the other (Woodward, 2003).

In addition to the above four forms of analysis pertaining to event causation, it is useful to consider for DSR the separate category of agent causation. Agent causation 'refers to the act of an agent (person, object) in bringing about a change' (Kim, 1999, p. 125). Thus, my flicking the light switch is the cause of the light turning on. It can be seen that agent causation analysis in general could be seen as reducible to manipulation event analysis. That is, the movement of my hand (an event) caused the light to come on (another event) and both these events were preceded by other events in a chain (walking through the door, perceiving that the room was dark). In this case, the act of an agent can be seen as *reactive*. It is a consequence of the agent's beliefs, attitudes and environmental inputs (Pearl, 2000).

Some have claimed, however, that one form of agent causation is not reducible to event analysis—namely, *substantival causation* (Kim, 1999). This form of causation is particularly relevant in design disciplines and we will distinguish it with a fifth form of causal analysis in our framework.

5. *Substantival causation analysis* (mental causation): This form of analysis recognises the creation of a novel or genuinely new substance or artefact by a human or humans, going beyond the mere change or manipulation of existing substances or their rearrangement. Many inventions would be examples of the effects of this type of causation—for example, the first telescope, the first bicycle and the first decision support system. The ability of humans to project virtual realities which do not yet exist (Ramiller, 2007) is a necessary but not sufficient cause in the design of artefacts. Recognising this type of causality requires recognition that humans have free will and can choose to do or create things that did not exist before, and these things themselves can play a part in other causal relationships. This type of causation recognises the *deliberative* (rather than reactive) behaviour of humans in exercising free choice (Pearl, 2000). The issue of the connection between the mental deliberations of humans and their consequent observable actions is part of a larger mind–body problem, which is beyond the scope of this essay. We will, however, distinguish this type of causation separately, because of its implications for design work.

Some further discussion of concepts relevant to causality is necessary to clarify some basic assumptions underlying the essay and our usage of terms. First, a cause is seen as an event or action that results in a change of some kind. If nothing changes then there is no cause and no consequent effect—that is, there is no change of state (Schopenhauer, 1974). Further, we have to consider the distinction between active causes and contextual causal conditions (which are more static). These each pertain to the issue of necessary and sufficient conditions, as these are central to many arguments for causality and to counterfactual analysis specifically. A counterfactual argument rests on the

claim that effect *E* would not have occurred if cause *C* had not occurred; in this case *C* is a *necessary cause* for *E*. To use a highly simplified example, the application of a burning match to a material could be seen as a necessary cause for a fire to light; however, there are other contextual conditions that are also needed for a material to ignite—for example, there must be enough oxygen present. Thus, though the match is necessary, it is not sufficient to cause a fire in the absence of other contributing contextual factors. But, taken together, the active cause and the causal condition (striking match plus oxygen) could be considered necessary and sufficient conditions for the fire to light. But even in this relatively simple case, there are problems in specifying all of the contextual conditions that are needed for both necessity and sufficiency. It might be that the active causal intervention of the burning match is not necessary, because some other active event could cause the fire to light (for example, lightning, spontaneous combustion, a spark from an electrical wiring fault). Further, it is difficult to specify all the contextual conditions that are necessary; in this case we have not specified that the ignited material must be flammable and it must be dry (for example, there must be an absence of water). The problem of complete determination of necessary and sufficient conditions verges on the impossible except for very simple, well-defined and closed systems.

It is for this reason that the words *ceteris paribus* (all else being equal) are added to claims to narrow the scope of the claim. For example, with our relatively simple case of the fire, we could claim 'given the existing state conditions (flammable material, oxygen, absence of water) and in the absence of other causes (spontaneous combustion, electrical spark), if the match had not been brought into close proximity to the material, the fire would not have started'. The causal claim is that close proximity of a burning match, *ceteris paribus*, is a necessary and sufficient condition for starting a fire in other (similar) situations. Here we see that causal claims are a form of generalisation (Lee and Baskerville, 2003) in which a theory specifying causal relationships is generalisable to other instances in similar contexts. In this way all design theories are claims that, *ceteris paribus*, an artefact built with the specified principles will cause the predicted outcome. Implied but rarely recognised in DSR is the necessary condition that the artefact be implemented and used successfully. As noted by Venable (2006), the problem space is composed of many related and potentially conflicting concepts, goals and stakeholders and the designed artefact will cause the predicted outcomes only if it is used in a manner consistent with the problem space as defined in the meta-requirements.

In socio-technical systems, however, we have to deal with situations where the number of causal conditions is large and there can be considerable uncertainty about the nature of linkages between cause and effect (Fay and Moon, 1996). Problem spaces in which artefacts will be implemented only rarely (if ever) fit

ceteris paribus conditions. In such situations it is useful to consider probabilistic reasoning about necessary and sufficient conditions. Pearl (2000) has advanced thinking in this area, and provides detailed coverage of how such reasoning can be dealt with for identification of causality using statistics. Pearl (2000, p. 284) shows how the 'probability of necessity' can be thought of in terms such as 'the probability that disease would not have occurred in the absence of exposure' to an infection. The disease might occur in only 1 per cent of cases without exposure. If you are not exposed you have a 99 per cent chance of not getting the disease; exposure is 'almost' a necessary condition. Similarly, the 'probability of sufficiency' can be expressed in terms such as the probability that a healthy, unexposed individual would have contracted the disease had he or she been exposed. The disease might follow exposure in 70 per cent of cases. There are links between this type of reasoning and the type of analysis that is needed in information systems. For example, the probability of necessity for module testing to ensure all errors are detected in programming is 99 per cent (1 per cent of cases would be error free if no module test occurs). The probability of necessity emphasises the absence of alternative causes that are capable of explaining the effect. The probability of sufficiency of a committed project champion is 80 per cent (80 per cent of cases with a committed project champion will be successful). The probability of sufficiency emphasises the presence of active causal processes that can produce the effect. The intricacies of determining necessary and sufficient conditions are laboured somewhat here because of their importance in reasoning in design science research. It is a very common form of analysis even if not recognised explicitly. Examples are cross-case analyses where an attempt is made to identify 'key' factors that are necessary, sufficient or both for some outcome to occur.

Some other aspects of causality are worthy of note for design disciplines. In the case of the fire lighting, a necessary condition is that the fire material is flammable. That is, the fire consumes fuel that is conducive to being lit. In information systems design fields, particularly in human–computer interaction, something like this notion is captured by the idea of 'affordance'. As explained by Norman (1988), the affordances of an object are the action possibilities that are readily perceivable by an actor because of the object's design characteristics. An example is a door that has no handle on the side that is to be pushed rather than pulled.

Another consideration is the causal characteristic of random interplay (Bunge, 2008), which results in emergent and unpredictable effects. Although these effects cannot be controlled or predicted, conditions that enable emergent behaviours and outcomes to arise from the lack of tightly coupled integration of components can be *designed for* in the evolution or secondary design of information systems (Germonprez et al., 2007). Systems in use consistently

show unexpected consequences (Dourish, 2006; Winograd and Flores, 1986), and Ciborra (2002, p. 44) notes that new systems of value emerge when users are 'able to recognise, in use, some idiosyncratic features that were ignored, devalued or simply unplanned'.

The concepts of both affordance and tailorable design are important because they are potential *causal conditions* that enable or constrain actions with the artefact that cannot be foreseen at the time of the design. They are potentially players in indeterminate chains of causal events. We will recognise the importance of this type of causality by distinguishing it as a sixth type of causal reasoning in design science research.

6. *Enabling causal condition* analysis: This analysis involves consideration of how artefact characteristics and conditions constrain or enable subsequent causal outcomes during use. The important characteristic is that the inclusion or exclusion of particular design characteristics will change the likelihood of the outcomes. This type of analysis is similar to type four, in which an intervention (an event or act) brings about an effect or makes an effect more likely. Here, however, we are separating active causes and contextual causal (enabling) conditions. In the example we gave previously, the act of striking a match was the active causal condition, whereas the placing of combustible material in the room by an agent was an enabling causal condition. A further example is perceived affordance, in which elements allow or encourage *possible* actions that are latent in the design. Examples include the scroll wheel on a computer mouse and roll-over text that informs users what will happen if they select a specific hyperlink. Another example of an enabling causal condition is the use of component architectures and recognisable conventions (Germonprez et al., 2007) that enable users to recognise conventional functions of component parts that can be reassembled into new patterns or adapted to new task functions. The design principles for the artefact are loosely coupled to the world so that users can create new structural couplings in alignment with their domain of action (Winograd and Flores, 1986). Design conventions such as icons that resemble functions performed (for example, a waste bin for 'delete', an hourglass for 'wait') guide people towards correct usage. This is a probabilistic cause in that most people familiar with icon conventions and symbols will understand the implied function and act accordingly.

The focus in design is often on obtaining a specific set of outcomes. But design can also include the goal of preventing specific outcomes (for example, preventing unauthorised system access, designing a 'rigid' artefact that users cannot modify in use). In this case the designer seeks to identify and eliminate necessary conditions for the undesired outcome or to find causes that obtain conditions that prevent the undesired outcome.

Note that we are discussing causality with reference to the work of a number of scholars who have made important contributions in this area. Our arguments have some congruence with ideas expressed in contemporary 'critical realism'—a philosophical approach based on work by Bhaskar (1975, 1998). There are, however, various schools of thought that could be termed critical realism and here we are providing an analysis of causality at a more fundamental level, relying on work that focuses specifically on this problem area.

A Framework for Causal Analysis in Design Science Research

Our analysis of causal mechanisms has pointed to six important perspectives for analysing causality that can potentially be used by researchers in design science research. Some of these perspectives are used commonly (if implicitly) by researchers and some are less common. But few design science researchers explicitly analyse and identify the causal claims upon which proposed design principles or theories are founded. In this section, we advance a framework that indicates how the different ways of thinking can be used to good effect.

First, we need to say something about the different types of artefacts that are dealt with, as the artefact type also influences the type of causal reasoning that is appropriate. Little discussion in the DSR literature distinguishes among the different classes of information systems produced. Recent research on design of organisations suggests a design distinction based on teleological goals. The same reasoning applies to the design of information systems intended to address different problem domains or different purposes. For example, the work of Germonprez et al. (2007) theorises about a class of artefacts that is intended to be modified in the context of use. This raises an interesting question regarding design theories for artefacts that are intended to be 'rigid' or explicitly not modifiable by the end user. In the first case, aspects of the system are emergent and therefore the *a-priori* causal analysis might be limited. In the second instance, the type of causal reasoning required will include causal analysis of how to prevent an outcome or how the absence of a feature might be a cause of something not occurring. A starting point for developing a framework for causal reasoning in DSR is a functional typology (Figure 3.1) of information systems such as that proposed by Iivari (2007), from which we can abstract dimensions for a causal framework.

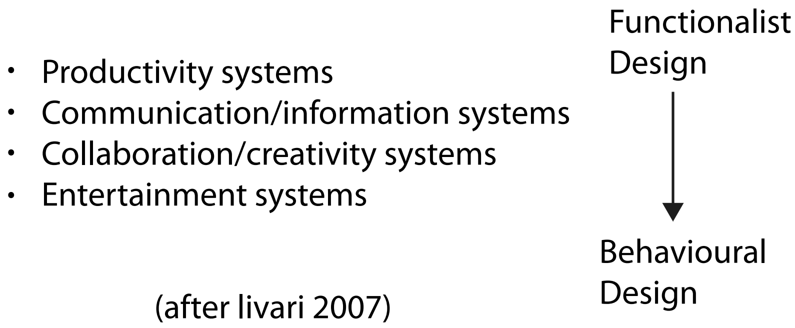

Figure 3.1 Teleological Abstraction of Information System Typology

This highly abstracted typology identifies a dimension along which all information systems fall. On one end are highly functionalist systems (Hirschheim and Klein, 1989) designed predominantly as productivity systems intended to achieve well-defined outputs with maximum efficiency from well-understood processes. As the processes, inputs, outputs and interactions are well known and understood, the causal connections and boundaries in the problem space are also well understood and the outcomes are highly predictable. Thus, specific types of causal reasoning are required. Characteristic of highly functionalist systems are tight coupling and strong component integration such as are found in accounting information systems and enterprise resource planning systems.

These systems can be contrasted with systems incorporating more behaviourally oriented design such as highly interactive collaboration systems or interactive entertainment systems, which privilege flexibility, creativity, adaptation to new problem domains and secondary design (Germonprez et al., 2007). This class represents design domains in which the users' behaviour and intentions are not only present, but are also required by the artefact-in-use. The contexts, tasks and users are diverse and variable and the systems are likely to evolve new patterns of in-situ use as they are modified. To obtain desired outcomes of system use requires types of causal reasoning that are enabling or probabilistic. Examples include design principles for learning systems or emergent knowledge processes (Markus et al., 2002). This distinction between design goals suggests a dimension of planned-emergent design, which forms one axis of our framework.

The other axis of our framework is formed by a distinction between the theorising that is done in designing artefacts (the interior prescriptive mode where artefacts are constructed to alleviate problems in the problem space) and the closely linked exterior descriptive mode, composed of the interactions of the artefact with its embedded context and its evaluation (Gregor, 2009). Although it is possible to conflate the interior and exterior modes of Gregor (2009) with the build and evaluate phases of Hevner et al. (2004), the distinction is important.

The interior mode focuses on theorising how artefacts can be designed and brought into being and is closely related to the build phase of Hevner et al. (2004). Here is where kernel theories are synthesised and causal connections are specified. Specifically, the abductive logic by which the explanation contained in the kernel theory (which is at a specific level of analysis and specific degree of generalisability) can explicitly define the connection between the principle and the expected outcome in the new design theory. The interior mode will often be iterative, with ongoing testing and experimentation helping to guide the design.

In contrast, the exterior mode focuses on the artefact-in-use after design is relatively complete or stable and the artefact is studied as part of a wider system, often by people other than the original designers. The exterior mode potentially includes all types of investigation, including measures of process or system output changes, user and management perception studies, phenomenological or hermeneutic studies of attached meaning and power structures or resistance. In this way, the exterior mode is differentiated from the evaluate phase of Hevner et al. (2004), which is predominantly focused on changes in efficiency, quality and efficacy. Knowledge gained from exterior mode research should include identifying causal connections for any research phenomenon related to the artefact-in-use. This might include negative outcomes, new problems or unexpected emergent behaviours that will inform the evaluation of the value of the artefact and, more importantly, inform future design activity.

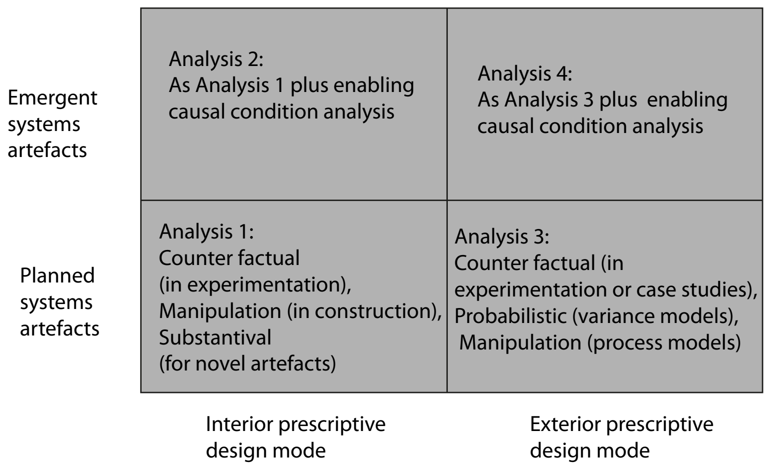

Figure 3.2 Types of Causal Analysis Useful in Design Science Research

Figure 3.2 shows the four cells that arise when these two dimensions are considered together, with indicative examples of appropriate causal reasoning given in each cell. The types of causal analysis suitable for each cell are now examined in more detail.

Analysis Cell 1: Interior design of planned systems

Examination shows that reasoning about causality in cells one and two differs in important ways from that in cells three and four, which are the cells associated with the traditional descriptive science approach. In these first two cells, the designer's thought processes in conceptualising a problem space and generating theoretical principles for potential solutions are themselves causal mechanisms. In the design of consequential management theory, Argyris (1996) suggests that the human mind functions as the designing system. This is what we term *substantival causality* (deliberative or mental causation). If we understood the direct causes or enabling conditions for human creativity and innovation, the design process could be manipulated to produce improved designs. But much design theory building is non-rational, abductive and unstructured. Reliance on kernel or reference theory to justify the 'idea' of the artefact is only part of the story; in many cases, we cannot say where the idea for the design came from, or why it is as it is, as human creativity and invention have come into play. Yet to evaluate the theoretical design principles and contribute to transferable knowledge, the design should be grounded in some type of reasoning that is amenable to causal analysis. To our knowledge, this type of causal analysis has not previously entered into discussion of design science research.

Many types of causal analysis can be used in both cell one and cell two, and DSR can be improved if they are explicitly applied. Manipulation analysis is used implicitly—that is, our team built this artefact and put it into use, with the implied prediction and expectation of a certain outcome. Here the analysis might consist simply of identifying what intervention will be created by the artefact and what system or behavioural change is expected as a direct result. This can be based on kernel theory, which demonstrates support for the causal linkage between manipulation and effect.

Counterfactual and probabilistic reasoning about causality are also used in an iterative design process. That is, the researcher constructs a prototype and experiments to see what results it causes, or does not cause, possibly in a probabilistic fashion. For instance, what percentage of test subjects prefer type A design to type B design? Iterative prototyping is inherently a process of refinement through identification of necessary and sufficient causal conditions. By adding or excluding specific physical conditions (affordances, conventions), psychological states (motivations, system explanations, user 'buy-in' through

participatory design) and goal modification (final cause), the designer searches the design space for the constellation of causal conditions that increases the probability of production of the desired effects.

An example is given in Codd's work on the relational database model (Codd, 1970, 1982). Codd made claims about how fewer mistakes would occur with use of relational databases because users would not have to expend so much effort dealing with the complexity of repeating groups. This is counterfactual analysis: the removal of the artefact feature of repeating groups from the human-use process is the cause of fewer errors.

Analysis Cell 2: Interior design of emergent systems

Although it seems counterintuitive to conjoin design and emergence there is a strong impetus to create some types of artefacts whose functions, applications and behaviours are flexible, agile and emergent. In addition to the types of analysis supporting cell one artefacts, there is the need to consider enabling casual condition analysis. In this type of analysis, specific design principles are selected because of evidence that they will increase the probability that a desired outcome will be encouraged or supported. As specific emergent phenomena cannot be predicted, the principles that will improve the likelihood that general desirable characteristics (for example, flexibility, mutability, ability to be reconfigured) will emerge are selected. These might be *conditional causes* where the designer considers enabling (or disabling) environmental conditions that increase the probability of an outcome (Sloman, 2005). Examples include identification of causes that are likely to create perceived affordances, secondary design or combinatorial application of functions (for example, services). Principles such as component architectures, recognisable conventions and metaphors (Germonprez et al., 2007) suggest necessary but not sufficient causal conditions for the potential of emergent system behaviour. Counterfactual analysis can be applied in reverse to identify factors or processes that rigidly couple system components to the world, resulting in brittle, inflexible system use (Winograd and Flores, 1986).

The design work by Braa et al. (2007) is an example of theorising in this cell. They call their work action research but they offer design principles. For example, to create a new health standard in a context that is characterised as a complex adaptive system, one should actively create an attractor—one of a limited range of possible states about which the system will stabilise. Another example is in service-oriented systems, in which the user creates relationships among services by determining types and relevancy of data and outputs, and what things go together (Hovorka and Germonprez, 2008).

Analysis Cell 3: Observation of planned systems in exterior mode

The reasoning about causality in this cell can employ the methods of counterfactual analysis advanced by authors such as Shadish et al. (2002) for experimental and quasi-experimental work. For example, claims for the advantages of the relational database model in terms of the hypothesised reduction in programmer error and greater ease of use could be tested in experiments. Case studies can also use counterfactual analysis in pattern analysis. We turn again to Braa et al. (2007) who examined cases of attempts to develop health standards in several different countries. They analysed chains of events (process models) in each case but they also contrasted what happened and did not happen in each country (a form of counterfactual analysis).

Probabilistic analysis can be done using statistics, in what is often referred to as testing of variance models, accompanied by reasoning about why causal effects should hold and how other explanations for effects can be ruled out. In many cases, however, the reasoning from statistical analysis relies on correlations and analysis of covariance. Researchers should be more aware of statistical techniques recommended for attribution of causality (see Pearl, 2000). Further, claims for causality can be examined in terms of manipulation analysis when process models are examined.

Analysis Cell 4: Observation of emergent systems in exterior mode

Attribution of causality in this situation is difficult precisely because the outcomes were not actually designed for, but rather emerged from the in-situ use of the artefact. Yet as Gregor and Jones (2007, p. 326) note, 'the ways in which [artefacts] emerge and evolve over time and how they become interdependent with socio-economic contexts and practices' are key unresolved issues for design. Numerous researchers have noted that artefacts are often used in ways that were not intended due to tinkering or secondary design of the system (Ciborra, 2002; Hovorka and Germonprez, 2010; Romme, 2003) and the inability of designers to share the same model of the design space as held by the users (Dourish, 2001). As noted in cell two, here, design principles to enable or constrain emergent system behaviours can be designed into the artefact, but particular emergent characteristics cannot be predicted.

In the evaluation of emergent system behaviours, probabilistic counterfactual analysis might be possible and even desirable. Determination of what causal mechanism was present that enabled emergent behaviours broadens the scope and fruitfulness of design theory. In other instances of emergent behaviours, the

design knowledge contribution might be in identifying mechanisms by which to extinguish or prevent behaviours. For example, secondary design of interfaces is not desirable in enterprise accounting systems or systems that require many information hand-offs. The principles for designing 'rigid' artefacts that are not amenable to secondary design are a largely unexplored area.

In concluding this section, we note that, not unexpectedly, in no cell was the first type of causal reasoning distinguished by Kim (1999) found to be relevant for socio-technical information systems. Because of the socio-technical complexity of designed and implemented information systems, we could find no example of causal reasoning that employed the logic of uniform and constant covering laws.[2] This observation has significant implications for the use of kernel theories empirically grounded in statistical evidence. As the kernel theories are only predictive in a probabilistic sense, derived design principles are frequently probabilistic. For DSR, this increases the knowledge creation burden on the evaluation phase, notably as counterfactual analysis can be used to identify the contexts or interactions in which the desired outcomes were not obtained.

Discussion and Conclusions

This essay has examined how causal reasoning can be employed in design science theorising. It has developed a framework with six types of potential causal analysis. The first four types are for event causation and include regularity analysis, counterfactual analysis, probabilistic analysis and manipulation analysis. A further two types are for agent causation and consist of substantival causation and enabling causal condition analysis.

Further, the essay develops a second framework that identifies the types of causal analysis that are suitable in different forms of DSR theorising. The orthogonal axes of this framework note distinctions on two different dimensions: 1) a planned versus emergent type of designed system; and 2) whether the work is in the interior prescriptive mode or the exterior descriptive mode of research. The four cells are labelled: 1) design of relatively stable planned systems; 2) design of emergent systems; 3) observation of planned systems in exterior mode; and 4) observation of emergent systems in exterior mode. The type of causal reasoning that can be used in each cell is described, with examples.

The question of substantival or mental causation in particular, although controversial, is worthy of attention because of its linkage to the truly novel

2 That the technical aspects of socio-technical systems are expected to behave in a uniform and predictable manner (for example, electronic circuitry) leads some researchers to reason in terms of covering laws. Cell one is where such reasoning, which we argue is very specific and limited, would appear.

artefacts that are a primary goal of design. Those reflecting on their research in DSR should consider how novel their artefact is. Genuinely novel and useful ideas and insights are likely to have greater impact. Codd's relational database work fell into this category. Reflection can distinguish novel innovations from new 'appliances' that might be more the result of normal industry practice, where knowledge of requirements plus knowledge of partial existing solutions that can be extended or adapted will cause an artefact to be produced in a fairly reactive fashion.

Our essay is significant because the topic of causal reasoning in DSR has received little, if any, attention. Our analysis has revealed ways of thinking about causality that have not been previously identified in the DSR literature. The position underlying the essay is that DSR can be better grounded by making clear the internal and theoretical warrants that underlie the theorising. Clarifying the causal claims invoked through kernel theories will improve theorising by providing criteria for kernel theory selection and delineating means of evaluation of the design theory based upon the assumed underlying causal claims. Even so, this essay also recognises that design theory can result from inspiration rather than theoretical or empirical grounding. But clear and explicit reasoning about causality and the different types of causal reasoning are a critical part of knowledge creation in the evaluation of design theorising. Causal reasoning has been shown to be an essential part of theory construction (Gregor, 2006). Our essay has practical implications because design theories underpin the construction of artefacts that are used in the real world, where the use of the artefacts can have consequences for both societal harm and societal good.

References

Abbott, A. (2001). *Time Matters: On theory and method*. Chicago, Ill.: University of Chicago Press.

Argyris, C. (1996). Actionable knowledge: design causality in the service of consequential theory. *The Journal of Applied Behavioral Science, 32*(4), 390–406.

Bhaskar, R. (1975). *A Realist Theory of Science*. London, UK: Verso.

Bhaskar, R. (1998). *The Possibility of Naturalism: A philosophical critique of the contemporary human sciences*. London, UK: Routledge.

Boland, R. J. (2002). Design in the punctuation of management action. In R. Boland (ed.), *Managing as Designing: Creating a vocabulary for management education and research*. Frontiers of Management Workshop, Weatherhead School of Management, 14–15 June.

Braa, J., Hanseth, O., Heywood, A., Mohammed, W., & Shaw, V. (2007). Developing health information systems in developing countries: the flexible standards strategy. *MIS Quarterly, 31*(2), 381–402.

Bunge, M. (1967). *Scientific Research II: The search for truth*. Berlin, Germany: Springer Verlag.

Bunge, M. (2008). *Causality and Modern Science*. [Fourth edn]. London, UK: Transaction Publishers.

Ciborra, C. (2002). *The Labyrinths of Information*. Oxford, UK: Oxford University Press.

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM, 13*(6), 377–87.

Codd, E. F. (1982). Relational database: a practical foundation for productivity. *Communications of the ACM, 25*(2), 109–17.

Collins, J., Hall, N., & Paul, L. A. (eds) (2004). *Causation and Counterfactuals*. Cambridge, Mass.: MIT Press.

Dourish, P. (2001). *Where the Action Is: The foundations of embodied interaction*. Cambridge, Mass.: MIT Press.

Dourish, P. (2006). Implication for design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*, Montreal, Canada, 541–50.

Fay, B. (1996). General laws and explaining human behavior. In M. M. Martin & L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (pp. 91–110). Cambridge, Mass.: MIT Press.

Fay, B., & Moon, B. (1996). What would an adequate philosophy of social science look like? In M. M. Martin & L. C. McIntyre (eds), *Readings in the Philosophy of Social Science* (pp. 21–35). Cambridge, Mass.: MIT Press.

Germonprez, M., Hovorka, D., & Callopy, F. (2007). A theory of tailorable technology design. *Journal of the Association for Information Systems, 8*(6), 351–67.

Goldkuhl, G. (2004). Design theories in information systems—a need for multi-grounding. *Journal of Information Theory and Application, 6*(2), 59–72.

Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly, 30*(3), 611–42.

Gregor, S. (2009). Building theory in the sciences of the artificial. *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology (DESRIST)*, Philadelphia, Pa.

Gregor, S., & Jones, D. (2007). The anatomy of a design theory. *Journal of the Association for Information Systems, 8*(5), 312–35.

Hevner, A. R., Ram, S., March, S. T., & Park, J. (2004). Design science in IS research. *MIS Quarterly, 28*(1), 75–106.

Hirschheim, R., & Klein, H. K. (1989). Four paradigms for information systems development. *Communication of the Association for Information Systems, 32*(10), 1199–216.

Hooker, J. N. (2004). Is design theory possible? *Journal of Information Technology Theory and Application, 6*(2), 73–82.

Hooker, R. (1996). *Aristotle: The four causes—Physics II.3*. Retrieved from <http://www.wsu.edu:8080/~dee/GREECE/4CAUSES.htm>

Hovorka, D., & Germonprez, M. (2008). Identification-interaction-innovation: a phenomenological basis for an information services view. *Information Systems Foundation: Answering the unanswered questions about design research*. Canberra, Australia.

Hovorka, D., & Germonprez, M. (2010). Reflecting, tinkering, and tailoring: implications for theories of information systems design. In H. Isomaki & S. Pekkola (eds), *Reframing Humans in Information Systems Development* (pp. 135–49). Berlin, Germany: Springer.

Hovorka, D. S., & Germonprez, M. (2009). Reflecting, tinkering, and bricolage: implications for theories of design. *Proceedings of the Fifteenth Americas Conference on Information Systems*, San Francisco, Calif.

Hume, D. [1748] (2004). An enquiry concerning human understanding. In L. P. Pojman (ed.), *Introduction to Philosophy: Classical and contemporary readings*. New York, NY: Oxford University Press.

Iivari, J. (2007). A paradigmatic analysis of information systems as a design science. *Scandinavian Journal of Information Systems, 19*(2), 39–64.

Kim, J. (1999). Causation. In R. Audi (ed.), *The Cambridge Dictionary of Philosophy* [Second edn] (pp. 125–7). Cambridge, UK: Cambridge University Press.

Lee, A., & Baskerville, R. L. (2003). Generalizing generalizability in information systems research. *Information Systems Research, 14*(3), 221–43.

Little, D. E. (1999). Philosophy of the social sciences. In R. Audi (ed.), *The Cambridge Dictionary of Philosophy* (pp. 704–6). Cambridge, UK: Cambridge University Press.

Markus, M. L., Majchrzak, A., & Gasser, L. (2002). A design theory for systems that support emergent knowledge processes. *MIS Quarterly, 26*(3), 179–212.

Mill, J. S. (1882). *System of Logic Ratiocinative and Inductive: Being a connected view of the principles of evidence and the method of scientific investigation*. New York, NY: Harper & Brothers.

Norman, D. (1988). *The Design of Everyday Things*. New York, NY: Basic Books.

Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge, UK: Cambridge University Press.

Ramiller, N. C. (2007). Virtualizing the virtual. In K. Crowston, S. Sieber & E. Wynn (eds), *Virtuality and Virtualization* (pp. 353–66). Boston, Mass.: Springer.

Romme, A. G. L. (2003). Making a difference: organization as design. *Organization Science, 14*(5), 558–73.

Salmon, W. (1998). *Causality and Explanation*. New York, NY: Oxford University Press.

Schopenhauer, A. (1974). *On the Fourfold Root of the Principle of Sufficient Reason*. La Salle, Ill.: Open Court Publishing Company.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, Mass.: Houghton Mifflin.

Sloman, S. (2005). *Causal Models: How people think about the world and its alternatives*. Oxford, UK: Oxford University Press.

van Aken, J. (2004). Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological rules. *Journal of Management Studies, 41*(2), 219–46.

Venable, J. (2006). The role of theory and theorising in design science research. *Proceedings of the First International Conference on Design Science Research in Information Systems and Technology*, Claremont, Calif., 1–18.

Winograd, T., & Flores, F. (1986). *Understanding Computers and Cognition: A new foundation for design*. Norwood, NJ: Ablex Publishing Corporation.

Woodward, J. (2003). *Making Things Happen: A theory of causal explanation*. New York, NY: Oxford University Press.