MANCHESTER 1824

CCSR

The Cathie Marsh Centre for Census and Survey Research

End Game: Can Game Theory Help Us Explain How a Statistical Disclosure Might Occur and Play Out?

CCSR Working Paper 2011-02
Elaine Mackey and Mark Elliot
Elaine.Mackey@manchester.ac.uk, Mark.Elliot@manchester.ac.uk

To fully account for the risk of a statistical disclosure occurring requires that we develop a better understanding of how a disclosure might occur and what its consequences might be. To do this we need to consider not just the target data but also the environment in which that data is produced and released. Through this we can identify and explore the events leading up to and following from a (claim of) disclosure. That is, we can move beyond modeling the mechanisms of statistical disclosure to conceptualising and systematically representing disclosure events in their entirety. In this paper, we show how it is possible to apply a game theoretic reasoning to model disclosure events to examine how key agents in a disclosure event might interact to bring about particular outcomes. The paper gives a brief introduction to game theory and to how it might be applied to the disclosure event situation.

www.ccsr.ac.uk

TRANSACTIONS ON DATA PRIVACY 1 (2011) 1002 —1050

End Game: Can Game Theory Help Us Explain How a Statistical Disclosure Might Occur and Play Out?

Elaine Mackey* and Mark Elliot**
* Centre for Census and Survey Research
University of Manchester
Oxford Road, Manchester M13 9PL UK
E-mail elaine.mackey@manchester.ac.uk
** Centre for Census and Survey Research
University of Manchester
Oxford Road, Manchester M13 9PL UK
E-mail mark.elliot@manchester.ac.uk

Abstract. To fully account for the risk of a statistical disclosure occurring requires that we develop a better understanding of how a disclosure might occur and what its consequences might be. To do this we need to consider not just the target data but also the environment in which that data is produced and released. Through this we can identify and explore the events leading up to and following from a (claim of) disclosure. That is, we can move beyond modeling the mechanisms of statistical disclosure to conceptualising and systematically representing disclosure events in their entirety. In this paper, we show how it is possible to apply a game theoretic reasoning to model disclosure events to examine how key agents in a disclosure event might interact to bring about particular outcomes. The paper gives a brief introduction to game theory and to how it might be applied to the disclosure event situation.

1 Introduction

The main concern of Statistical Disclosure Control (SDC) is to ensure a balance is reached between confidentiality and data utility (for various perspectives on this issue see: Duncan et al 2001; Kennikel and Lane 2006; Purdam and Elliot

2007; Shlomo 2007; Trottini 2003). To make decisions about how best to optimise this trade off, we must first assess the risk of a disclosure occurring and this, in turn, requires that we establish an understanding of the processes and events that might lead to a disclosure. We posit that a fully constituted understanding of statistical disclosure processes, such as *intrusion*, will only be achieved by answering each of three questions:

1. what are they?'
2. how might they happen?
3. what would their consequences be?

To date, the SDC literature has focused heavily on the mechanisms of disclosure i.e. they have been addressing the 'what they are?' question with little attention paid to the other two. This is an unusually restricted approach to risk research which typically considers precursors and payloads as well as mechanisms. Thus there is a pressing need to strengthen our understanding of the 'precursors' and 'consequences' of the disclosure process. However, this is far from straightforward because data is disseminated in a complex *data environment* of which we have only partial knowledge. This situation is further compounded because SDC research takes as its object 'data' (most usually just the data to be released) rather than the data environment which consists of data, systems and most critically agents. So, in order to improve our understanding of the precursors and consequences of disclosure we have to shift our focus away from the data towards human agency, because it is only when we look at the actions of, and interactions between, the key agents in the data dissemination setting that we can begin to understand how disclosure events might be created, their outcomes shaped and their likely impact felt.

2 Disclosure Processes and Events

Let us begin by defining the terms *disclosure processes* and *disclosure events*. A disclosure event is an occurrence arising from an interaction between agents[1],

[1] Agents in the data dissemination setting include also data users, the general public, media, specialist interest groups and SDC specialists. Although these agents might at first glance appear to be on the periphery, they have a key role in helping us understanding how a disclosure event might play out. These agents might be drawn into a direct interaction with the DSO as a primary player or indirectly as a secondary player.

typically (but not necessarily) including a data stewardship organisation (DSO)[2] and an intruder, which leads to the perception of disclosure and/or the claim of such. It consists of a set of precursors or enabling conditions, a disclosure process and a set of consequences. A disclosure process is one whereby a linkage (veridical or otherwise) between a data unit and a population unit is created. Thus a disclosure event may or may not be based on an actual disclosure and it may or may not be made public.

The point at which action commences, with the DSO making decisions about a potential data release, to when an agent believes he has uncovered a plausible disclosure/disclosure claim is the narrative for a disclosure event. We can create many different narratives of disclosure for each disclosure process by exploring the actions of key agents and the interactions between them. In line with the terminology used in SDC we shall refer to theses narratives as scenarios from here on in.

The use of scenarios as a tool for developing a better understanding of how a disclosure might occur is not new in SDC. The work of Paass (1988) Mokken et al (1992) Lambert (1993) and Elliot and Dale (1999) established scenario analysis as a key part of disclosure risk assessment. The usefulness of scenarios lies in their potential to illuminate a range of conditions required to bring about a potential disclosure. However, until now they have been used in a uni-dimensional and static way and though this provided a useful analysis of the possible actions of a single agent (referred to here as an intruder), it has not, crucially, allowed for an examination of the inter-action that occurs between key agents when, for example, a DSO releases data and a data user or would be intruder accesses it. An analysis of such interactions between agents is fundamental to our understanding of how a disclosure might actually arise and then play out. This is because, in this setting as in any other, the actions of all agents are interdependent i.e. each agent's actions affect and are affected by the actions of one or more other agents. Thus, to understand the probable actions of an intruder we must consider them in relation to the actions of the DSO. Similarly, to consider the best actions for a DSO we need to consider them in relation to the probable actions of an intruder (and responses of other key agents such as the media and the General Public). We might therefore hypothesise that the best actions of the DSO should be based on its perceptions of the best actions of the intruder. Similarly, the best actions of the intruder are likely to be based on his perceptions (and partial knowledge) of the best actions of the DSO. Of course, this means that the DSO should direct its efforts to avoid a disclosure by attempting to thwart the efforts of the intruder whilst the intruder is likely to di-

[2] After Duncan et al (2010) we use data stewardship organisation to refer to any organisation that has responsibilities for data dissemination processes.

rect all his efforts towards counteracting those of the DSO in an attempt to bring about a plausible disclosure. However, it also means that potentially the DSO (and the intruder) could also attend to the consequences of disclosure by making moves which alter those consequences. By focusing on multi-agent action and interactions in this way we can:

i. Identify who the key agents are in the data dissemination setting.
ii. Set out the possible actions agents are likely to employ when interacting with other agents.
iii. Examine how particular sets of agents' actions may lead to particular outcomes.
iv. Explore the consequences of these outcomes.

If we can establish points i-iv the potential benefits to DSOs will be considerable. First, once we have an understanding of the precursors and consequences of disclosure we can combine that understanding with what we already know about the mechanisms of disclosure to develop more accurate disclosure risk metrics. Second, establishing points i-iv gives us a basis from which to develop strategies aimed at constraining the actions of some types of agents and/or limiting the consequences of those actions. In short, over and above its potential to help us better, understand and, assess risk it potentially offers us an additional means for managing the disclosure problem.

We can identify four types of disclosure process for which one could develop disclosure event scenarios: actual intrusion, simulated intrusion, spontaneous recognition and spontaneous disclosure. Here, in order to simplify our discussion, we shall refer only to the disclosure processes of intrusion (actual and simulated). Further, we will be focused on aggregated data releases only. Nevertheless, the principles described here could equally well be applied to microdata releases and to other forms of disclosure processes[3]

The precise form of any given disclosure event scenario will be shaped by the DSO's data dissemination decisions regarding: the type of data released, the SDC methodology employed, the level of detail given and the publication of metadata, as well as other key agents' counter-responses to these decisions. For example, a DSO releases a dataset (the makeup of which it determines) which is then accessed by a would-be intruder. The intruder has, we argue, (at least) two types of response that he could employ. The intruder's success in applying these

[3] Indeed although in general we are focussed on statistical disclosure here the same framework could applied to any data security event (e.g. data losses or hacker attacks).

responses will not just be influenced by the DSO's data dissemination policies but also his own skills, knowledge, access to resources and motivations. The intruder's possible responses are: type 1, exploit the released data using other (external) sources of data[4] to pursue an actual disclosure; type 2, manipulate public perceptions about what constitutes a disclosure to pursue a simulated disclosure claim. For response option 2 to have any potential for success it would require that the DSO's data release strategy be potentially vulnerable to manipulation such as the release of tables of counts containing zeros (which may or may not be true counts) giving the impression that new information could potentially be revealed or the release of tables of counts containing uniques giving the impression that a respondent could be identified. The key issue here is that in using this response type the intruder aims to make others such as the media or general public believe that a disclosure is possible. This response strategy is likely to be most effective when a combination of factors come into play: at times when a DSO provides very limited information about the issue of data confidentiality and when there is low public confidence in their (or similar organisations) handling of confidential data.

So far we have considered the case for examining the interactions between agents rather than focusing on the actions of a single agent (the intruder). Further, we have looked descriptively at a single interaction between the DSO and intruder. If we wish however to move beyond a descriptive examination to consider how agents' interactions might play out, and what the consequences of particular outcomes might be, we need a framework for modeling one agent's action in relation to another agent's action. One such framework is Game Theory.

3 Game theory[5]

Given the interdependent nature of the relationship between agents in the data dissemination setting, a game theoretic framework could, we suggest, be applied to the interactions between our agents. More precisely, it can help us to develop plausible disclosure scenarios through which to explore the likely interactions between key agents as well as the outcomes and consequences of those interactions.

[4] The obtaining of external data by the intruder may be done using legitimate or non-legitimate means.

[5] In producing this summary of game theory we have drawn on several general texts on the topic. Specifically: Dixit And Nalebuff 1991; Binmore 2007; Straffin 1993; Camerer 2001; 2003; Wright 2002; McCain 1997.

As Levine (2000) says, game theory can be described as a 'science of strategy'. It uses mathematics to predict the actions that an agent[6] will or, given certain assumptions about motives, preferences, should take, in an interaction, to obtain the best outcome whilst accounting for the possible actions of other agents.[7] Within game theory each agent has a strategy profile i.e. a list of all his possible actions (strategy choices). Each agent's strategy choices in a game, of which there may be many, are commonly represented by numerical payoffs based on their preferences. These preferences are ordered and mapped onto real numbers, via a utility function. For example, suppose agent one prefers A to B and B to C, the ordinal[8] utility function maps the highest ranked preference with the largest number and then the next highest ranked preference with the next largest number and so on. This may be represented like this: A >> 3, B >> 2, C>>1. The aim of each player is to end the game with the largest payoff. This, however, will only be partially within each agent's control since the outcome, and therefore payoffs, will depend upon the strategy choices of all agents in the game. Agents' behavior is described as *strategic* because they take into account this interdependence when they choose which action to take in the game.

In general, it is also assumed that players are rational with the specific meaning that they play to maximize their utility. The notion of rationality is a core concept in game theory as it underpins both descriptive and predictive analysis about how a game should/will be played. It is thought that if one can work out the best response of each player to the actions of his opponent then one can find the game's (unique) solution[9]. However, rational play involves an individual

[6] An agent (or player) can be virtually any entity: a person, a nation or business (Camerer 2001).

[7] Game theory does not attempt to explain why agents act in the way that they do although this does not mean that game theorists are not concerned with analytical type questions. They are, it is just that these types of questions are directed at the game level. For example, why do certain game conditions produce particular outcomes?

[8] An ordinal utility function tells us about the ranking of a player's preference; it does not inform us about the strength of his preference. To account for an agent's strength of preference we would require a cardinal utility function. However, it is questionable whether one can apply a cardinal utility to non-economic outcomes (see Fang et al 1993 for discussion).

[9] Solution concepts can be thought of as rules that predict the actions of players. There are several solution concepts such as Von Neumann and Morgenstern 'Minimax Theo-

making complicated decisions about: (i) how to choose a strategy with a desirable outcome for oneself whilst knowing that one's opponent is also trying to choose a desirable outcome and (ii) decisions about whether one's interests coincide or conflict with one's opponent. Not unsurprisingly, in practice, players do not always act to maximize their utility. As Straffin points out: 'In the real world, it is doubtful that all players will play rationally', (1993:4).

The idea that players don't always play rationally raises something of a problem. This is because, if the game's outcome is dependent on the choices each player makes, the presence of a player that does not play to ensure his preference is never violated changes what a rational opponent should do. There have been attempts to take account of a less stringent model of rationality but so far it has had limited success. Nevertheless, it is clear that definition of rationality as isomorphic with utility maximisation is itself equivocal and at the very least requires a very broad conceptualisation of utility to achieve applicable functionality. However, reflecting the fact that most game theory work has been done within economics, it is generally employed in a quite restricted sense of material equivalence. Here our starting position is to not restrict ourselves to any particular definition, but to propose that most of the agents in disclosure event games will be operating in a goal directed or motive lead fashion and we will assume that they will choose actions that they perceive move them towards those goals and/or are concordant with those motives.

4 A framework for analyzing disclosure event games

Rather than setting out an example of one of the very many games[10] that may arise from the data dissemination setting we shall provide an overview of the basic structure of such games. Thus we present no more than a general framework and a number of exemplars to illustrate it. The branch of game theory that underpins the game structure described here is *non-cooperative* game theory. Contrary to what one might imagine from its title, it does encompass interactions where there is a basis for cooperation as well as ones where there is not[11].

rem' for zero sum games and for non-constant sum games John Nash's Nash Equilibrium.

[10] For further discussion on game playing in the data dissemination setting see Mackey (2009).

[11] For example, the DSO and the intruder will have very obvious opposing interests with regards to data confidentiality: the DSO wishes to protect data confidentiality and the intruder wishes to breach it. But they may also have interests that coincide, for example

The crucial element is that when agents do decide to co-operate this cooperation is self-enforcing. There are two types of games that are of interest to us they are: (i) coordination games where players share a mutual interest in playing the same strategy and (ii) anti-coordination games where players share a mutual interest in playing different strategies.

Let us begin our description of a game in the data dissemination setting by considering a simple interaction involving a DSO and an intruder. In this interaction, we shall hypothesise that, the DSO makes the first move by setting the initial conditions of the game and then releasing data into the public domain. Let us suppose that the intruder, in this simple interaction, responds by either making a single move or two consecutive moves. The intruder's first move involves him making an assessment of the likelihood of his success, in respect of his goals, and selecting (what he believes to be) an optimum strategy from the following:

(i) to do nothing and therefore abort the disclosure attempt and end the interaction.

(ii) to pursue a simulated attack by manipulating public perceptions.

(iii) to pursue an actual attack by fishing for a named individual.

(iv) to pursue an actual attack by linking the disseminated data to an external data source.

After choosing options ii, iii or iv, the intruder goes onto make a second move: to makes his attempt known to others or not.

The DSOs second move, after releasing data, involves it also selecting a strategy from:

(i) to make a pre-emptive move.

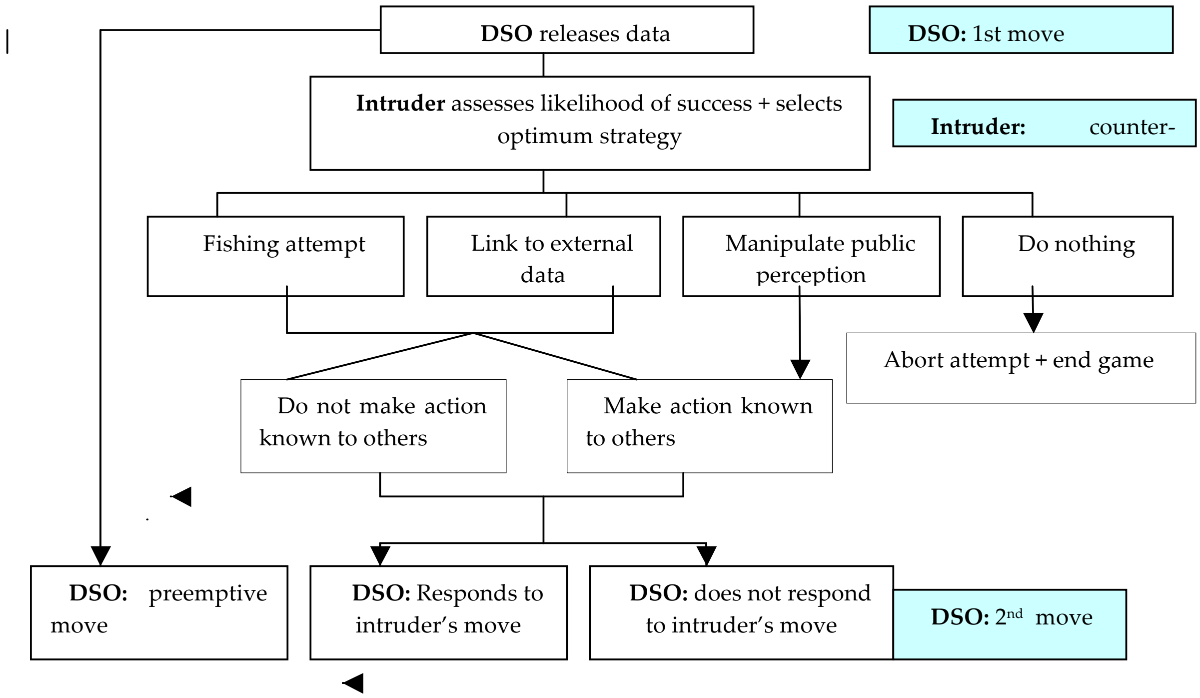
(ii) to do nothing.

(iii) to respond to the intruder's move.

This series of moves is illustrated in figure 1.

to keep a disclosure out of the public domain. For the intruder this may be because he wishes to avoid detection and for the DSO this will because it wishes to avoid disrupting the status quo of its relationship with respondents and the wider public.

Figure 1 DSO-Intruder Interaction



To illustrate further the core elements of game playing we shall use this very basic interaction to set out a number of hypothetical games. All games start with a condition of play. Here, this is established in the data dissemination processes that the DSO utilises when releasing a dataset. The intruder makes the next move aware of the move that has gone before. In figure 1, the intruder's move involves him making decisions about whether or not to pursue an attempt and if so how best to respond to the DSO efforts to maintain data confidentiality. This particular combination of sequential moves can be modeled as a game, represented in extensive form by a directed graph or game tree, to illustrate an order of play. In this 'disclosure attempt game' framework we can vary the moves. For example, rather than a single second move there might be simultaneous moves by the intruder and the DSO. With, for example, the intruder deciding whether to employ optimum strategies choices i, ii, iii or iv at the same time as the DSO deciding to make pre-emptive move by delivering a publicity campaign around the issues/problems of maintaining data confidentiality. Here, the DSO's move is called *pre-emptive* because it is not aware of the intruder's move whilst it is choosing its move. We might imagine that it plays this move because it is considering a long term public relations strategy aimed at both building a rapport with the General Public and ameliorating the potential damage to its reputation in the event of a public disclosure claim being made. In this form of the game the players are not aware of *all* the moves in the game. It can be represented in strategic form using a bi-matrix table which lists the payoffs of the players' strategy choices but no order of play. In simple sequential games the order of play can influence the outcome of a game. However, in more complex social games such as those in the data dissemination setting it is unlikely to be this simple. An order of play will only be one factor that may influence the outcome of a game. In games where the players' interests are opposed, one player holding a credible threat will be another such factor, whilst in games where players have a mutual interest in playing the same strategy information abstracted from outside the game[12] (such as a player's knowledge of their opponents preference etc) may be an influencing factor.

If we look at figure 1 again, we see that there is a further game framework we could explore which we might refer to as the 'disclosure claim game'. In this game, we start from the position that an attempt has been made and the intruder believes he has grounds to make a plausible disclosure claim. This game can be modeled as a set of sequential moves involving the intruder, first, deciding whether or not to make a public disclosure claim and the DSO deciding whether or not to respond. We shall return to explore this game more fully in a moment.

So far we have described a basic form of the interaction between DSO and intruder and identified the types of games (sequential or simultaneous) that might be played by these players but how do we play these games? In other words, and this is one of the key issue for applying Game Theory to the data dissemination setting, how do we determine (plausible) strategy choices for our players and their preference for one strategy choice over another?

[12] This is Schelling's (1960) theory of focal points, which theorises how players may coordinate on a particular equilibrium (or rather game outcome) when they cannot communicate.

4.1 A method for Applying Game Theory

As we have said above, a central requirement for applying Game Theory is the development of a method for formally: (i) identifying plausible actions (strategy choices) and (ii) exploring preferences for one action over another.

In terms of identifying plausible actions for our players, we suggest that one should begin by asking the question '*what does a player hope to achieve?*' in any particular disclosure event scenario. To illustrate this with an exemplar based on the disclosure claim game, let us imagine that the intruder is thinking about whether or not to make his disclosure claim known to others; so, what might his motivations be? Let us say that he is motivated by one of three things, to: (i) upset the status quo by for example embarrassing the DSO; (ii) obtain information, (iii) challenge him-self i.e. his attempt is about proving to himself or maybe others that he can breach the DSO data security measures. We can use this list of possible motivations as a basis for determining his list of actions. We might hypothesise that his list of possible actions are: (i) make a public disclosure claim; (ii) establish a dialogue with the DSO; (iii) make no declaration of his activities.

The DSO's list of possible actions can be similarly determined by considering it's motivations[13] in responding to the intruder. The key question once again is about '*what it hopes to achieve?*' in any particular disclosure event scenario. Let us say that it is motivated by one of two things, i.e. to: (i) establish public confidence in its actions; (ii) prevent a disclosure becoming public knowledge.[14] Given the DSO's list of motivations we might hypothesise that its list of possible actions are: (i) establish a dialogue with the intruder; (ii) establish a public dialogue with other agents about the intruder's claim; (iii) make no public or private declaration on the matter of the intruder's activities; (iv) make no public declaration on the matter of the intruders action but threaten/ pursue a prosecution. Having established the list of possible actions for the players in the disclosure claim game the next step is to determine which action they might prefer, which we will now illustrate.

To help determine a player's preference for a particular action from those available to him/it we consider the notion of *approaches or roles* a player may take in an interaction, which would lead them to favor one action over another. We shall consider (for simplicity of exposition) two approaches for each of our players although many more will be possible. For the intruder, let us suppose

[13] For expository purposes, we will put aside here the ontological question of whether organisations can be said to have motivations.

[14] We acknowledge that these are not mutually exclusive (nor indeed are the intruder's own motivations) but for expository purposes in the current context we are treating them as if they are independent.

he could take one or other of the following approaches related to his goals: *covert* and *overt*. Thus we describe his approach as covert if his objective is to keep his actions hidden and overt if he wants to make his actions known. For the DSO, let us suppose it could take one or other of the following approaches relating to its goals: *engaging* and *reserved.* Thus we describe the DSO's approach as reserved if its objective is to avoid being drawn into a public discussion on disclosure and engaging if the reverse is true.

Assigning the players different approaches enables us to both establish a more formal way of addressing the issue of players' preference and explore multiple variations of a single game. So for the approaches outlined above we have four possible variations of the disclosure claim game:

(1) The intruder takes a covert approach whilst the DSO approach is reserved.

(2) The intruder takes an overt approach whilst the DSO approach is reserved.

(3) The intruder takes an overt approach whilst the DSO approach is engaging.

(4) The intruder takes a covert approach whilst the DSO approach is engaging.

Taking game 2 as an example, we see that the intruder favors an overt approach whilst the DSO favors a reserved approach. So how might we model the players' actions using this combination of preferred approaches as a framework? The first thing to do is to set out the player's list of possible actions in order of preference. Thus, we hypothesise that if the intruder takes an overt approach then he will prefer strategy choice (i): *make a public disclosure claim* to strategy choice (ii): *establish a dialogue with DSO* and this to strategy choice (iii): *make no declaration of his activity*. If we were to assign a numerical payoff to his list of possible actions (as outline in the Game Theory section of the paper) we would write it thus:

Make public a disclosure claim >> I. Ai3

Establish a dialogue with DSO>> I. Aii2

Make no public declaration of disclosure attempt activity >> I. Aiii1

Key: I denote the intruder.

Ai, Aii, Aiii, represents the Intruder's list of possible actions.

Number 1, 2, 3, represents the payoff associated with a particular action.

Similarly, we might hypothesise that if the DSO takes a reserved approach then it will prefer strategy choice (i): *make no public or private declaration on the matter of the intruder's claim* to strategy choice (ii): *make no public declaration on the matter but pursue a private prosecution* to strategy choice (iii): *establish a dialogue with the intruder* and this to strategy choice (iv): *establish a dialogue with other* *agents about the intruder's claim*. Thus we can assign numerical payoffs to the DSO list of possible actions as follows:

Make no public or private declaration on the matter>> D. Ai4

Makes no public declaration on the matter but pursue a prosecution >>D.Aii3

Establish a dialogue with the intruder>> D. Aiii2

Establish a dialogue with other agents about claim>> D. Aiv1

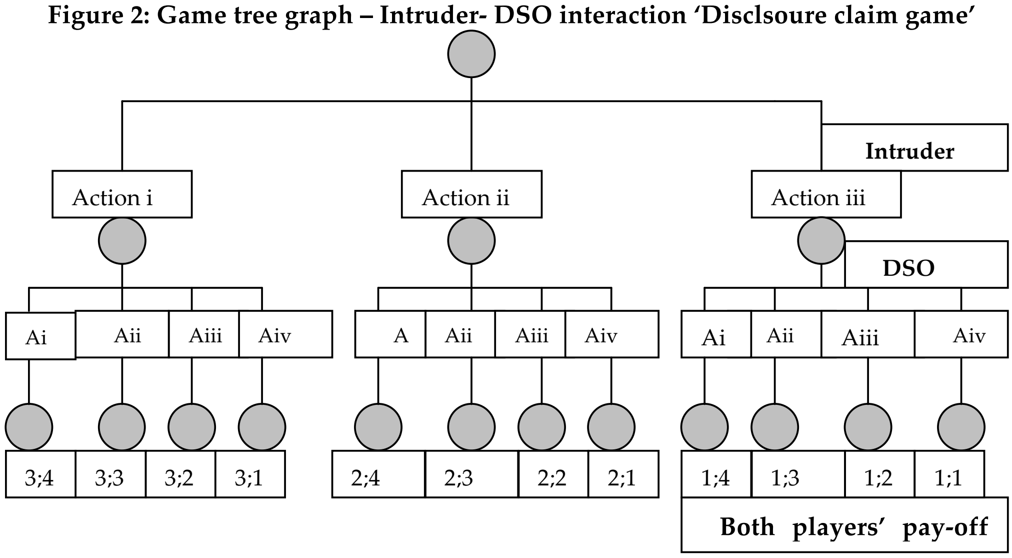
Key: D denotes the DSO

Ai, Aii, Aiii, Aiv, represents the DSO list of possible actions

Number 1, 2, 3, 4 represents the payoff associated with a particular action

This game can be set out using a game tree graph as it is charaterised by a sequence of moves. Figure 2 illustrates this. To summarise, the game has the following features: (i) a list of players, i.e. the DSO and intruder, (ii) a list of possible actions for each player, (iii) players who know each other's list of possible actions, (iv) players who prefer one outcome over another and (v) players who act to maximise their payoffs.

Figure 2: Game tree graph – Intruder- DSO interaction 'Disclsoure claim game'



Looking at the graph, how might we work out what the outcome of the game is likely to be? The key issue here, as we have discussed previously, is that a player makes decisions about which of his actions to play based on obtaining a desirable outcome for himself whilst knowing that his opponent is also trying to obtain a desirable outcome. The interdependent nature of the interaction means that our players will be drawn into recursive reasoning:

'The intruder thinks that the DSO thinks that the intruder thinks that the DSO thinks ……………'.

Game theory uses a small number of *solution concepts* for the purpose of short circuiting this recursive reasoning.

One such solution concept is the Nash Equilibrium (NE). An NE is said to occur where there is a set of actions with the property that no player can benefit by changing his action unilaterally. Thus the question we are interested in is what is the best cumulative payoff for our players? Let us explore this for the disclosure claim game by imaging it plays out thus:

(i) The intruder, lets call him Bob, would like to play his strategy with the highest payoff i.e. *to make public his disclosure claim*. However, Bob reasons that if he plays this strategy it is highly likely that the DSO will pursue a prosecution, which may lead to sanctions that he would rather avoid. Thus he concludes that his next best action in line with his overt approach is to: *establish a dialogue with the DSO*. He reasons that by playing this action he can still upset the status quo at the same time as avoiding a prosecution because the DSO will not want others to know about his claim as even a private prosecution runs the risk of being made public.

(ii) The DSO is aware of Bob's move prior to making its own. It reasons that to pursue a private prosecution in response to Bob's move could lead to others finding out about Bob's claim. It further reasons that to do nothing meaning that it does not engage with Bob, could lead to Bob making a further move to make his claim public. Thus it concludes that its best response in line with its reserved approach is to engage with Bob and: *establish a dialogue with him.*

The game's Nash Equilibrium then is: the intruder plays *establish a dialogue with the DSO* and the DSO plays *establish a dialogue with the intruder* [Aii2; Aiii2]. As even this simple example illustrates, the complexities of modeling moves involve reference to motivational structures, and reasoning about the intention of others. A full analysis, involving all possible agents, would be very complex with agents operating at various scales from individual psychological though to organizational and beyond. Nevertheless pursuit of such a *grand model* will enable us to better understand the precursors and consequences of disclosure events – and this is *our* primary motivation in pursuing this line of research.

5 Concluding remarks

In this paper we have outlined the case for expanding the scope of disclosure risk assessment beyond the mechanics of disclosure to an analysis of the environment in which disclosure events take place. We have outlined how a game theoretic reasoning can help us understand better how disclosure events might arise and be played out in the data dissemination setting. Another advantage of the approach outlined here is that it has wider application to data types than the official microdata and tabular aggregates that are the focus of orthodox disclosure risk analyses. Confidentiality issues surrounding social network data, administrative datasets and qualitative data would all be amenable to treatment using game theoretic techniques. The possibility of encapsulating a larger range of data types and data confidentiality issues in a single general framework is very attractive. We feel that the approach has something significant to offer the field and will be pursuing it further in future work.

To fully realise the potential of game theory's application to SDC (to develop a more empirically grounded method for assessing disclosure risk and thereby a more principled approach to addressing the competing demands of data confidentiality and utility) there are several areas which need development. We have presented here a framework for simple two player games. One key area for development is a detailed study of more complex, multiple move (and multiple player) games. So far we have imagined that the intruder and DSO engage in a single isolated interaction. Our players' actions will not however occur in a void, indeed this is what we recognize and explicitly explore when we consider the notion of consequences. Thus our next move will be to build on the structural aspects of game playing set out in this paper.

We envisage that our future research will fall into three categories: (i) pre-game issues (ii) game issues and (iii) post-game issues. In terms of pre-game issues in addition to developing the structural aspects of game playing we shall further develop the framework outlined in the paper for identifying players' actions and preferences. In terms of game issues, we will need to consider how the form of rationality that we propose here might be refined and to what extent it needs to be modified to encompass institutional as well as individual decision making. Finally, in terms of post-game issues we will need to develop further a framework for interpreting game playing and in particular around the concept of credible threat and the balance of power between players.

Acknowledgements

The research reported here was carried out supported by funding from the Economic and Social Research council and the Office of National statistics under CASE award S422134054.

References

[1] Doe, J., Smith, R. (2005) The paper example number one. In *Proceedings of the best conference in the world (BCITW 2005)*, pages 99-110. ACM

[2] Author, F., Author, S., Author, T. (2008) Title of the paper: On Author's family, Transactions on Data Privacy 1 1-2.

[3] This is the 2nd reference, and we are using the "Reference" style.

Binmore, K. (2007). Game Theory: A Very Short Introduction. Oxford University Press.

Binmore, K. (1994). Game Theory and Social Contract: Playing Fair. MIT Press.

Camerer, C. F. (2003). *Behavioural Game Theory: Experiments in Strategic Interaction*. Princeton University Press.

Camerer, C. F. (2003). Behavioural Studies of strategic thinking in games. *TRENDS in Cognitive Sciences*. Vol.7. No.5. May 2003. Elsevier. pp225-235.

Camerer, C. F.; Teck-Hua Ho.; Chong, J.K. (2001). *Behavioural Game Theory: Thinking, Learning and Teaching*. http://www

Crawford, V. (1997). Theory and Experiments in the Analysis o Strategic Interactions in Kreps, D. and Wallis, K. (eds). *Advances in Economics and Econometrics: Theory and Applications.* Seventh World Congress. Vol1. Econometrics Society Monographs No 27. Cambridge, New York. Cambridge University Press. Pp206-242.

Dickson, E. and Nowaczyk, P. (2004). *Introduction to Game Theory in Political Science*. Online at http://www.nyu.edu/gsas/dept/politics/undergrad/syllabi/v53.0840-dickson-/04.pdf.

Dixit, A. and Nalebuf, B. (2002). *Game Theory*. The library of Economics and Liberty. The Concise Encyclopaedia of Economics. Online at *http://www.econlib.org/library/Enc/Gametheory.html*.

Duncan, G. T., Keller-McNulty, S. A., and Stokes, S. L. (2001): "Disclosure Risk vs. Data Utility: The R-U Confidentiality Map", Los Alamos National Laboratory Technical Report LA-UR-01-6428.

Elliot, M.J. and Dale, A. (1999). Scenarios of Attack: The Data Intruder's Perspective on Statistical Disclosure Risk. *Netherlands Official Statistics*. Special edition, Spring, pp 6-10.

Elliot, M.J. (1996). *Attacks on census confidentiality using the Sample of Anonymised Records*. Presented at the Third International Seminar on Statistical Confidentiality, Bled, Slovenia, October.

Fiori, S. (2005). Simon's Bounded Rationality. Origins and use in Economic Theory. Working Paper No. 09/2005

Fudenberg, D. and Levine, D. K. (1998). *The Theory of Learning in Games*. MIT Press.

Fudenberg, D. and Tirole, J. (1991). *Game Theory*. MIT Press.

Kennickell, A. B. and Lane, J. (2006): "Measuring the impact of data protection techniques on data utility: Evidence from the Survey of Consumer Finances", in J. Domingo-Ferrer, ed. *Privacy in Statistical Databases* (Lecture Notes in Computer Science):291-303.New York: Springer-Verlag.

Lambert, D. (1993). Measures of Disclosure Risk and Harm. *Journal of Official Statistics*. Vol. 9, No. 2, pp 313-331.

Levine, D. (2000). *An overview of Game Theory*. Online at http://www.levine.sscnet.ucla.edu/general.htm, **Error! Hyperlink reference not valid.**, accessed 9.02.2007.

Mackey, E (2009) A Framework for Understanding Statistical Disclosure Processes, PhD Thesis, University of Manchester.

Marsh, C., Skinner, C., Arber, S., Penhale, B., Openshaw, S., Hobcraft, J., Lievesley, D. and Walford, N. (1991). The Case for Samples of Anonymised Records from the 1991 Census. *Journal of the Royal Statistical Society* 154(2) pp 305-340.

McCain, R. (1997). *Strategy and Conflict: An introductory Sketch of Game Theory*. Online at http://william-king.www.drexel.edu/top/eco/game/game.html, accessed 10/01/20008.