Reading agendas between the lines, an exercise

Sileno, G.; Boer, A.; van Engers, T.

CrossMark

Reading agendas between the lines, an exercise

Giovanni Sileno[1] · Alexander Boer[1] ·
Tom van Engers[1]

Abstract This work presents elements for an alternative operationalization of monitoring and diagnosis of *multi-agent systems*, developed in the context of compliance checking. In contrast to traditional accounts of model-based diagnosis, and most proposals concerning non-compliance, our method does not consider any commitment towards the individual unit of agency. Identity is considered to be mostly an attribute to assign responsibility, and not as the only referent to a source of intentionality. The proposed method requires as input a set of prototypical *agent-roles* known to be relevant for the domain, and an *observation*, i.e. evidence collected by a monitor agent. We elaborate on a concrete example concerning tax frauds in real-estate transactions.

1 Introduction

In previous works (Boer and van Engers 2011a, b), we have presented a *model-based diagnosis* view on complex social systems like the ones in which public administrations operate. The general framework targeted by our research is intended to support administrative organizations in improving responsiveness and adaptability, for instance by enabling the streamlining of use cases and scenarios of non-compliance in the design cycle and in operations. This paper focuses in particular on the *operationalization* of model-based diagnosis (to be used in operations, and therefore supporting responsiveness) and differs from the previous papers in granularity, as it provides a specific example of implementation. Note that even if

✉ Giovanni Sileno
g.sileno@uva.nl

1   Leibniz Center for Law, University of Amsterdam, Amsterdam, Netherlands

we apply the proposed method to identify the occurrence of non-compliance, it may be used in principle for any other pattern that may be of interest for the organization.

The paper is organized as follows. Section 1 provides a general introduction to diagnosis, and to what we intend as diagnosis of social systems; Sect. 2 presents an overview on the various literature in AI about model-based diagnosis; Sect. 3 introduces the case study (sale transactions of real-estates), identifying prototypical scenarios of interest; Sect. 4 concerns the actual exercise of operationalization of monitoring and diagnosis, providing insights and directions for future developments.

2 Diagnosis of social systems

In general, a diagnostic process is triggered if there is the presumption that a *failure* occurred in the system. However, what counts as a failure depends on the nature and function of system.

In case of a *designed artifact*, the system is generally associated with a set of requirements, and, at least at the moment of production, to an implementation model—a *blue-print*. A *failure* becomes manifest when there is an inconsistency between the form/behaviour that is observed and what is expected from that artifact. The failure may be at the *design level*, when the implementation does not meet the design requirements; or at the *operational level*, when one of the sub-components has failed, and propagated its failure to the system.

In case of a *social system* (natural or artificial), the internal mechanisms of social participants are unknown and typically inaccessible. For instance, we are not able to fully know what is in the mind of a person, nor how someone's mind actually works (not even our own).[1] Nevertheless, we still *do* apply (when it is relevant to do so) a *theory of mind* to explain and interpret our own or others' behaviour, by referring to notions as beliefs, desires, and intentions. If we assume that the application of this stance is viable, then, when something goes wrong in a social system, i.e. when someone's expectations about the behaviour of someone else are not met, this means that something went wrong at an informational, motivational, or deliberative level of at least one individual.[2] In order to identify the wrong, however, we have to consider the requirements associated with the system.

A first filter of discrimination could be obtained by referring to normative directives: prohibitions and obligations correspond respectively to negative and positive requirements. This would be sufficient, if the contextualization of a generic norm in an actual social setting were straightforward. However, as the existence of the legal system shows, this is far from being the case: the *qualification* of actions, conditions and people and the *applicability* of rules build up the core of the *matter*

---

[1] In the words of Chief Justice Brian (1478): "for the devil himself knows not the thought of man".

[2] This is also true in domains where the law imputes *strict liability*, i.e. where the claimant only needs to prove the occurrence of the *tort*, and not of a *fault* (negligence, or unlawful intent) in the agent who performed the tort. In these cases, the law discourages reckless behaviour, pushing the potential defendant to take all possible precautions. In other words, in strict liability, law ascribes fault *by default* to the agents making a tort.

of law debated in courts. Thus, in an *operational setting*, rather than norms, we need to refer to adequate abstractions of cases, making explicit factors and their legal interpretation; in this way, we handle *contextualized normative models* that can be directly used to discriminate correct from faulty behaviour, all while maintaining a legal pluralistic view.[3]

2.1 Deconstructing identity

Current approaches of diagnosis on *multi-agent systems* (MAS) consider social system components (software agents, robots, or persons) as individual intentional entities, i.e. following an assumption that could be described as "*one body, one mind*" (see references in Sect. 2.1). In contrast, we assume that intentional entities may transcend the individual instances of the agents. In the case of a *combine* (e.g. in sport, when a player makes an agreement with a bidder on the results of a match) or similar schemes, the collective intentional entity that causes and explains the resulting behaviour is placed behind the observable identities. Such an interpretation of intentionality has a relationship with the notions of coordination, coalition formation, and distributed cognition.[4] In addition to this "*one mind, many bodies*" scenario, we allow that an agent may interleave actions derived by a certain strategy with actions generated for other intents, independent from the first: the "*one body, many minds*" case may apply as well.

2.2 Diagnosis as part of a dual process

Monitoring agents (e.g. tax administrations) are typically continuously presented with a stream of messages (e.g. property transfer declarations) autonomously generated by social participants. Clearly, they would encounter a cognitive overload if they attempted to reconstruct all "stories" behind such messages.

In affinity with Dual Process theories of reasoning, we may distinguish a *shallower*, less expensive but also less accurate mechanism to filter the incoming messages; and a *deeper*, more expensive, and accurate mechanism to analyze the filtered messages, possibly performing further investigative actions. The first, implemented as a *monitoring* task, is designed by settling what is interesting to be monitored, and which are the threshold conditions that identify *alarming* situations. The second, implemented as a *diagnostic* task, is triggered when such (potentially) alarming situations are recognized, and may start specific courses of actions to look for other clues discriminating possible explanations (diagnostic and non-diagnostic). Note that the two tasks are intimately related: they are both constructed using expectations of how things should go, and of how things may go wrong. Furthermore, planning builds upon abilities, which can be reinterpreted as expectations of how things may go performing certain actions in certain conditions.

---

[3] This may be useful for practical purposes: a public administration may, for instance, use dissent opinions of relevant cases to further strengthen its service implementations.

[4] cf. Hutchins (2010): "A central claim of the distributed cognition framework is that the proper unit of analysis for cognition should not be set *a priori*, but should be responsive to the nature of the phenomena under study."

From a practical reasoning point of view, *planning, monitoring and diagnosis are parts functional to a whole, and the practical reasoning of an agency cannot but be unbalanced if one of these functions is neglected.* This implies that all effort that a public administration puts into simplifying the operations in the front-office of service provision (e.g. diminishing the evidential burden on the citizen) should be coupled with effort in the back-office in support of institutional maintenance.

2.3 Side effects

The choice of investigative actions requires some attention as well. In the case of physical systems, *measurements* do not necessarily involve a relevant modification of the studied system (at least at a macro-level), and criteria in deciding amongst alternative measuring methods generally concern costs vs opportunities. In the case of a social system, this cannot be the only criterion. For instance, if the target component suspects being under observation, he may adopt an *adversarial* or a *diversionary behaviour* protecting him from intention recognition actions (cf. Sadri (2012)); he may also drop the unlawful intent as a precaution. In this work, we overlook the planning problem for evidence-gathering tasks that take into account these derived behavioural patterns.

3 Relevant literature

Model-based diagnosis is a traditional branch of study of AI (see e.g. Lucas (1998) for an overview); it has reached maturity in the 1990s, and it has been applied with success in many domains, reaching a production level of technology readiness (Console and Dressier 1999). In the following, we retrace the main directions of investigation, highlighting where relevant the specificities of our problem domain.

Consistency-based diagnosis Early approaches in model-based diagnosis used explicit fault models to identify failure modes (Davis 1984), but these were replaced by diagnostic systems based on descriptions of correct behaviour only. Practical reasons explain this progress: in the case of electronic devices, manufacturers provide only descriptions of normal, correct behaviour of their components. Failure modes could be computed simply as inconsistencies with the nominal specifications ((Reiter 1987) for a minimal set of faulty components, (de Kleer and Williams 1987) for multiple-faults configurations). This type of diagnosis is usually called *consistency-based diagnosis*. In short, by having models of correct behaviour of the system components and a topological model of their composition and knowing the initial state, we can predict the expected system state via simple deduction. If the observed output is different, we acknowledge a behavioural discrepancy, which triggers the diagnostic process aiming to identify the faulty components. Note that in this case, such components are deemed *faulty* simply because they do not behave according to their nominal specification: the 'negative' characterization is then constructed in duality to the 'positive' one (cf. *negation as failure*). In recent literature, these are also called *weak fault models* (WFM) (Stern et al. 2014). This

approach entails important consequences: in consistency-based diagnosis, all fault models become equivalent, meaning that, from the diagnoser perspective, "a light bulb is equally likely to burn out as to become permanently lit (even if electrically disconnected)" (de Kleer and Williams 1989).

Abductive diagnosis Not surprisingly, the approach provided by consistency-based diagnosis is not appropriate for certain domains. In medicine, for instance, doctors study not only the normal physiology of human organisms, but also how certain symptoms are associated with diseases; the hypotheses obtained through diagnosis are used particularly to *explain* given symptoms. In other words, 'negative' characterizations—*strong fault models* (SFM)—are asserted in addition to the 'positive' ones (cf. *strong negation*), rather than in duality to them. In the literature, in order to operationalize this approach, several authors have worked on explicitly characterizing the system with faulty models, starting a line of research which led to the definition of (model-based) *abductive diagnosis* (Cox and Pietrzykowski 1986; Console et al. 1989).

Type of diagnosis per type of domain We can sketch two explanations of why certain domains refer to consistency-based diagnosis, and others to the abductive diagnosis. The first explanation is built upon the use of negation. The former approach takes a *closed-world assumption* (CWA) towards the system domain, while the latter considers an *open-world assumption* (OWA), reflecting the strength of knowledge and of control that the diagnoser assumes having. Reasonably, engineering domains prefer the former (everything that does not work as expected is an error), while natural and humanistic domains usually refer to the latter (there may be a *justification* for why things didn't go as expected). The second explanation considers the different practical function for which diagnosis is used in the domain. While by applying consistency-based diagnosis we can identify (minimal) sets of components which are deemed to be faulty and that can be substituted for *repair*, in the second type of diagnosis the underlying goal is to diagnose the 'disease' in order to provide the right *remedy* (that often cannot be a substitution). For these reasons, considering the social system domain, it makes sense to deal not only with positive, normal institutional models (e.g. buyer and seller in a sale contract), but also with explicitly faulty ones (e.g. tax evaders).

Despite these differences, however, abductive diagnosis and consistency-based diagnosis have been recognized as two poles of a spectrum of types of diagnosis (Console and Torasso 1991). In effect, we find contributions extending consistency-based diagnosis with faulty models (de Kleer and Williams 1989) and abductive diagnosis with models of correct behaviour. In a more principled way, Preist et al. (1994) show that the two types of diagnosis can be unified relying on a *stable model semantics*—the same used in *anwser set programming* (ASP)—essentially because it considers the distinction and separate treatment of *strong negation* and *negation as failure*.

Selecting additional investigations During a diagnostic process, it is normal to consider the possibility of conducting additional investigations (measurements, in the case of electronic devices) in order to conclusively isolate the set of faulty components, or more generally, to reduce the set of hypothetical explanations. For simplicity, we will neglect this aspect in this work; for completeness, however, we

highlight two main directions investigated in the literature. The most frequently used approach, first proposed in (de Kleer and Williams 1989), is to use a *minimum entropy* method to select which measurement to do next: choosing the datum which minimizes the entropy of the candidate after the measurement is equivalent to deciding the source that provides the maximum *information* to the diagnoser (Shannon 1948). As this method considers only one additional source per step, it is also called *myopic*. The second approach proposes instead *non-myopic* or *lookahead* methods, i.e. deciding multiple steps to be performed at once (Heckerman et al. 1993). In principle, this is the way to proceed when we intend to minimize or control side-effects when deciding on strategies for collecting information.

3.1 Diagnosis of multi-agent systems

The association of diagnosis with *multi-agent systems* (MAS) is not very common in the literature, although the number of studies is increasing. In general, contributions consider only one of the two use cases of MAS, i.e. as a mechanism for distributed computation or as a framework for the instantiation of agent-based models. Therefore, on one side, MAS are proposed as a solution to perform diagnosis of (generally non-agent) systems, like in (Roos et al. 2003; Pipattana-somporn et al. 2009). On the other side, understanding of social failures is expressed as a problem of social coordination—see for instance (Kalech 2012; Kafali and Torroni 2012). Unfortunately, the latter have generally a design-oriented approach, and non-compliance and social failures are therefore seen has a design issue, rather than as systemic phenomena, as they would be in a "natural" social system. For this reason, they share a perspective similar to works on checking non-compliance at regulatory level, e.g. (Governatori 2013; Jiang et al. 2014): system (normative) requirements are literally taken as the reference against which to test compliance of business processes. Unfortunately, in doing this, we are not able to scope behaviours that superficially look compliant, but, for those who know the 'game', are not.

Using agent-roles instead of roles The idea of using normative sources is related to the *role* construct; agents are usually seen as enacting certain institutional/ organizational roles (Dastani et al. 2005), inheriting their normative characterization. An alternative approach, from which this contribution develops, has been proposed by Boer and van Engers (2011b), based on *agent-role* models: constructs that include the coordination of roles. The agent-role model shares elements with those used in *intention-recognition* studies, and in particular with those based on logic approaches—see Sadri (2012) for an overview—which have grown out of traditional AI accounts of story understanding and abduction. However, from a conceptual point of view, the "first principles" we are considering with agent-roles are not simple rules, but knowledge structures building upon practical reasoning constructs (Sileno et al. 2015) and institutional positions (Sileno et al. 2014c). More importantly, agent-roles are defined not only by a *script*, but also by a *topology*. By allowing multiple identities distributed on the topology, the agent-role model takes

into account the existence of *collective agencies*, transcending the individual social participants.

4 Case study: *swap schemes* in real-estate transactions

In the following section, we will focus on a well-known type of real-estate fraud, of the family of *swap schemes*, and present a few similar prototypical patterns. In a market context, a swap scheme establishes coordinations between dual groupings of buyers and sellers; as these parties are expected to compete within that institutional framework, it essentially undermines the *arm's length* principle of the market. On small economic scale this is not forbidden: e.g. "if you make me pay less for the guitar that your father is selling, I would make you pay less for my brother's motorcycle." However, in real-estate transactions, property transfer taxes apply. The full interaction includes the tax administration, and in these conditions swap schemes become means to reduce the amount of taxes due and, therefore, are not permitted.

4.1 Outline of a database of scenarios

Let us consider a simplified real estate market, with economic actors buying and selling houses of type A and of type B. Property transfer tax is 6% of the sale price, and the buyer and the seller have both nominally the burden to pay it (the actual distribution amongst the parties is however not fixed a priori). Besides the normal sale, we take into account three different scenarios: a swap scheme implementing a real-estate fraud, a hidden payment, and a wrong appraisal.

Example 1 (Real estate fraud, Swap scheme)X and Y want to exchange their properties: X owns a real estate of type A; Y owns one of type B, both worth €10 million. Instead of paying €600,000 per each in taxes, they set up reciprocal sales with a nominal price of €5 million, thus dividing the taxes due in half.

The scheme is illustrated in Fig. 1. The picture highlights two coordination levels:

- an *intentional coordination* level, generally referring to some composition of institutional roles (in our case buyer/seller structures, the dashed boxes in the figure);
- a *scenario coordination* level, responsible of the synchronization of operations between the intentional coordination structures.

The first is the domain of *internal topologies* of agent-roles. The second is the domain of *coupling* configurations of agent-roles, i.e. of *external topologies*, specified as MAS.
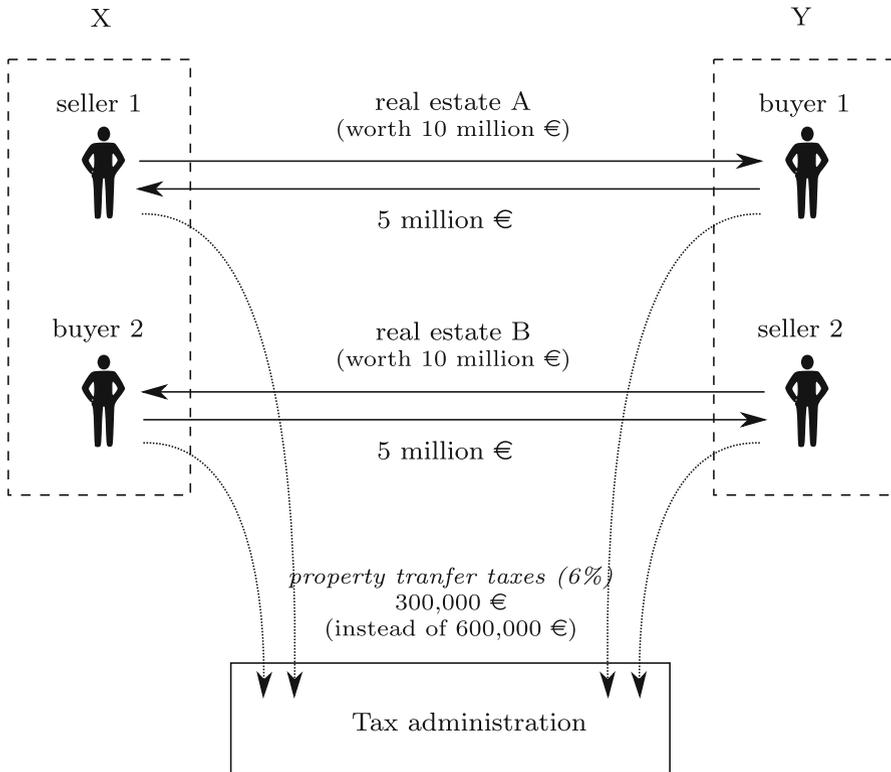
Fig. 1 Topology of a real estate fraud based on a swap scheme

The structures enabling coordination (at both levels) may be physical bodies, but also social bodies as natural, informal groupings of people (e.g. father and son), organizations (e.g. employer and employee), etc. It may be anything that suggests a sharing, a *concentration of interests*, or an existence of *stable inter-dependencies*, that may undermine the arm's length principle. At the scenario level, however, the relation is not necessarily as structured as the examples just given. In the case of bribery, for instance, there is typically no other relation between the parties beside a *contingent agreement*. Similarly, a swap-scheme may be performed by two real-estate agencies on a contingent basis.

Example 2 (Hidden payment) X wants to give €300,000 to Y, and, as Y is also interested in X's house, X sells Y that house, worth €500,000, for €200,000.

A hidden payment is usually economically advantageous for both parties because property transfer generally has lower taxation than other forms of transfer.

Example 3 (Wrong appraisal) X needs to sell his house. Not knowing the current prices for the area, he sells the house for €200,000 to Y, while at market price, the building would be worth around €500,000.

5 Operationalization of monitoring and diagnosis

In this exercise, we imagine taking the role of the tax administration, with the intent of monitoring the payment of taxes, possibly diagnosing (and also explaining) supposed institutional failures.[5] Note that the tax administration has only a *partial view* of the communications of the parties: in our simplified world, only sale declarations and tax payment receipts.

   Types of failures

   The starting point of the operationalization is to collect the agent-roles of the domain relevant to the tax administration. The first set is given by simple intentional characterizations of *normal institutional roles*, i.e. buyers and sellers paying their taxes. From this, we can construct possible failure modes as violations of role obligations, dealing with representations of *negative events* (negative as they are defined by the failure of expectations concerning events). In this specific example, tax payment may be:

(i)   completely missing, as failure to pay *tout court*,
(ii) wrong, as failure to pay the fixed amount of taxes (e.g. 6% of the sale price)
(iii) wrong, as failure to pay the 'right' amount of taxes, in terms of *reasonableness*, i.e. of what could have been expected to be paid to the tax administration for the sale of that property.

The third situation covers the case of swap-schemes or other tax evasion manœuvers; it is evidently more difficult to track down, as it requires an evaluation in terms of the social domain semantics—in this case, of the market pricing rationality. This is the domain in which the agent-role concept makes a crucial difference.

5.1 Monitoring

As we know that certain social participants may be non-compliant, we need to set up an adequate monitoring procedure. A first requirement of adequacy is the possibility of *discriminating* cases of non-compliance from those of compliance. This actually supports a general principle for choosing monitoring targets:

Proposition 1 *Outputs of contrast operations between compliant and non-compliant scenarios lead to identifying events or threshold conditions associated with suspicious transactions.*

   The set of discriminating elements is constructed in terms of what is available through the monitoring, i.e. the 'perceptual' system of the agency. If the diagnostic agent is not able to monitor any discriminatory element, then the contrasting principle will not be exploitable and there will be no mean to recognize non-

---

[5] It is worth observing that compliance and non-compliance are qualifications relative to the position of the diagnostic agent in the social system. For instance, in a world of liars, truth-tellers would fail with respect to the social practice of systematically lying.

compliance. In our example, as the tax administration has direct access only to sale declarations and tax payment receipts, it is amongst these sources that we have to scope signs of potential failures.

Note that the contrast operation can be implemented thanks to the availability of executable models: by *executing* normal and failure models, we can predict the different traces they would produce, and then contrast them. In principle, however, we could refer directly to the traces. For instance, in medicine, failure modes are usually directly associated with symptoms, without explaining why a certain disease produces these symptoms. In the general case, however, this solution has limitations, as it assumes a relative invariance of the chain of *transmission* going from the source phenomenon to the perceptual system of the observer, which is not granted in a social system. Considering explicitly the underlying behavioural mechanism allows us to deal separately with such 'transmission' component.

We apply the previous principle to the three types of negative events. Case (i) requires the implementation of a *timeout* mechanism that asynchronously triggers the failure. Case (ii) requires a check *synchronously* to the receipt of payment; it can be implemented with a simple operational rule. Case (iii) is more complex: to conclude that a price is reasonable requires us to assess the market price of that property, and to decide what deviation from market price is still acceptable. Let us arbitrarily specify this deviation as 40% of the market price, knowing that statistical methods may suggest more appropriate values. Therefore, the price provided in the sale declaration can be taken as a threshold to consider a certain sale price as *suspicious*. If implemented in Prolog, the qualification rule would look like the following code:

```
suspiciousPrice(Price, Estate, Time) :-
  marketPrice(MarketPrice, Estate, Time),
  Price =< (MarketPrice * 60)/100.

suspiciousSale(Seller, Buyer, Estate, Price, Time) :-
  declaration(sale(Seller, Buyer, Estate, Price, Time)),
  suspiciousPrice(Price, Estate, Time).
```

Clearly, this is a simple case. In general, multiple factors may concur with different weight to increase the suspiciousness of transaction.

In absence of average market price As we confirmed from talking with experts of the tax administration, the practical discrimination used by investigators to discover potential tax frauds is actually built upon comparisons with average market prices. Unfortunately, average market prices are not easy to be access in reality and, when they are, they may be not representative for that specific case.[6] A first solution

---

[6] On the one hand, prices of real estate properties in public offers often do not correspond to the actual prices of sale. On the other hand, the heterogeneity of real estate parcels, the imperfect alignment between cadastral information and real situations, the dynamics of value associated with neighbourhoods and other relevant factors make it difficult to consider as reliable the application of average measures on actual cases.

would then be to refer to domain experts, e.g. appraisal agents, but these externalizations, where available, obviously increase the costs of investigation. A simple way to overcome the problem of assessing the market price of a certain real-estate property is to check the value of the same real-estate in previous sale transactions. In the case of swap schemes, the new owners tend to sell the recently acquired property after a relatively short time, but for a much higher price, even in the presence of relatively stable prices. From an operational point of view, this would correspond simply to a different tracking of the suspiciousness relation.

5.1.1 Diagnosis

When identified, suspicious transactions should trigger a diagnostic process in order to establish *why* the failure occurred. In general, the same 'symptoms' may be associated with diagnostic and non-diagnostic explanations. For instance, going through the known scenarios, a low price in a sale transaction may be due not only to a swap scheme, but also to a hidden payment, or it may simply be due to an error in the appraisal of the estate by the offeror. Interestingly, even if plausible, wrong appraisal is not taken into account by the tax administration. Why? Evidently, this choice is determined by the *strict liability* of these matters[7], but it may be seen as a consequence of a more fundamental issue: the tax administration cannot possibly read the mind of offeror to check the veracity of his declaration. A price that is not 'reasonable' cannot but be interpreted as an *escamotage* of both parties to avoid or reduce the tax burden.

  Direct diagnostic mechanism In a simplistic form, direct evidence for a supposed swap-scheme would consist of two sets of buyers and sellers that have performed suspicious sales:

```
actionEvidenceOfSwap(
   sale(Seller1, Buyer1, EstateA, PriceA, Time1),
   sale(Seller2, Buyer2, EstateB, PriceB, Time2)
) :-
   suspiciousSale(Seller1, Buyer1, EstateA, PriceA, Time1),
   suspiciousSale(Seller2, Buyer2, EstateB, PriceB, Time2),
   not(EstateA = EstateB),
   not(Seller1 = Seller2), not(Buyer1 = Buyer2).
```

This is however not sufficient: sellers and buyers may have performed these transactions independently, and therefore this evaluation doesn't consider minimal

---

[7] See note 2.

circumstantial elements to support a swap-scheme rather than e.g. two hidden payments. In order to overcome this problem, we have to take into account explicitly a *relatedness* condition.

```
actionAndCircumstantialEvidenceOfSwap(
  sale(Seller1, Buyer1, EstateA, PriceA, Time1),
  sale(Seller2, Buyer2, EstateB, PriceB, Time2)
) :-
  actionEvidenceOfSwap(
    sale(Seller1, Buyer1, EstateA, PriceA, Time1),
    sale(Seller2, Buyer2, EstateB, PriceB, Time2)
  ),
  relatedTo(Seller1, SharedStructure1),
  relatedTo(Buyer2, SharedStructure1),
  relatedTo(Seller2, SharedStructure2),
  relatedTo(Buyer1, SharedStructure2).
```

An example of *relatedness* condition between buyer and seller may be, for instance, their participation in a common social structure (family, company, etc.), that may place its members outside the arm's length principle of the market. This condition acknowledges *potential* intentional coordination, i.e. a plausible concentration of *interests* that makes the transaction definitively suspect.[8]

The existence of a coordination structure at the scenario level, i.e. between such shared structures, would be additional evidence, but it is not necessary, as the scheme may be performed on a contingent basis Sect. 3.1. Interestingly, the 'hidden payment' case turns out to be a minimal version of a swap-scheme:

```
actionAndCircumstantialEvidenceOfHiddenPayment(
  sale(Seller, Buyer, Estate, Price, Time)
) :-
  suspiciousSale(Seller, Buyer, Estate, Price, Time),
  relatedTo(Seller, SharedStructure),
  relatedTo(Buyer, SharedStructure).
```

By extension, we could imagine swap-schemes implemented through *networks* of buyer and sellers. This would be, for instance, a simple diagnostic test for swap-schemes performed on three-node networks:

---

[8] This is evidently similar to the issue of *conflict of interest*: a person in power may be in a situation in which his discretion to reach the primary intents defined by his role may be biased towards the achievement of other intents.

```
actionAndCircumstantialEvidenceOf3Swap(
  sale(Seller1, Buyer1, EstateA, PriceA, Time1),
  sale(Seller2, Buyer2, EstateB, PriceB, Time2))
  sale(Seller3, Buyer3, EstateC, PriceC, Time3)
) :-
  suspiciousSale(Seller1, Buyer1, Estate1, PriceA, Time1),
  suspiciousSale(Seller2, Buyer2, Estate2, PriceB, Time2),
  suspiciousSale(Seller3, Buyer3, Estate3, PriceC, Time3),
  not(EstateA = EstateB),
  not(Seller1 = Seller2), not(Buyer1 = Buyer2),
  not(EstateB = EstateC),
  not(Seller2 = Seller3), not(Buyer2 = Buyer3),
  not(EstateA = EstateC),
  not(Seller1 = Seller3), not(Buyer1 = Buyer3),
  relatedTo(Seller1, SharedStructure1),
  relatedTo(Buyer3, SharedStructure1),
  relatedTo(Seller2, SharedStructure2),
  relatedTo(Buyer1, SharedStructure2),
  relatedTo(Seller3, SharedStructure3),
  relatedTo(Buyer2, SharedStructure3).
```

The inclusion of a third element breaks the direct connection between the initial parties, but the code makes explicit the pattern that can be extended by induction. More formally:

Definition 1 (*Generalized swap-scheme through sales*) Given n sale transactions, naming b_i and s_i respectively the buyer and the seller of a transaction i, a swap scheme holds if the following relatedness relations are established:

– between s_1 and b_n (named X_0)
– with $0 < i \leq n$, between s_i and b_{i-1} (named X_i)

The associated topology is illustrated in Fig. 2. It would certainly be interesting to evaluate mechanisms like this on data sets such as those released with the so-called Panama papers.

5.2 Improving the reasoning mechanism

The diagnostic mechanism proposed here leverages the advantages of *backward chaining* given by Prolog, i.e. of reasoning opportunistically in order to reach a conclusion about a certain epistemic goal. In a way, this is an opposite solution to the operationalization we proposed in *explanation-based argumentation* (EBA) (Sileno et al. 2014b), based on ASP, where factors brought by the observation are used to allocate *all* possible scenarios. The present proposal suffers from three important limitations. First, it relies on a *closed-world assumption* (CWA), i.e. *negation as failure* is automatically interpreted as *strong negation*. Second, it requires an explicit query to trigger the inferential process, but, in practice,
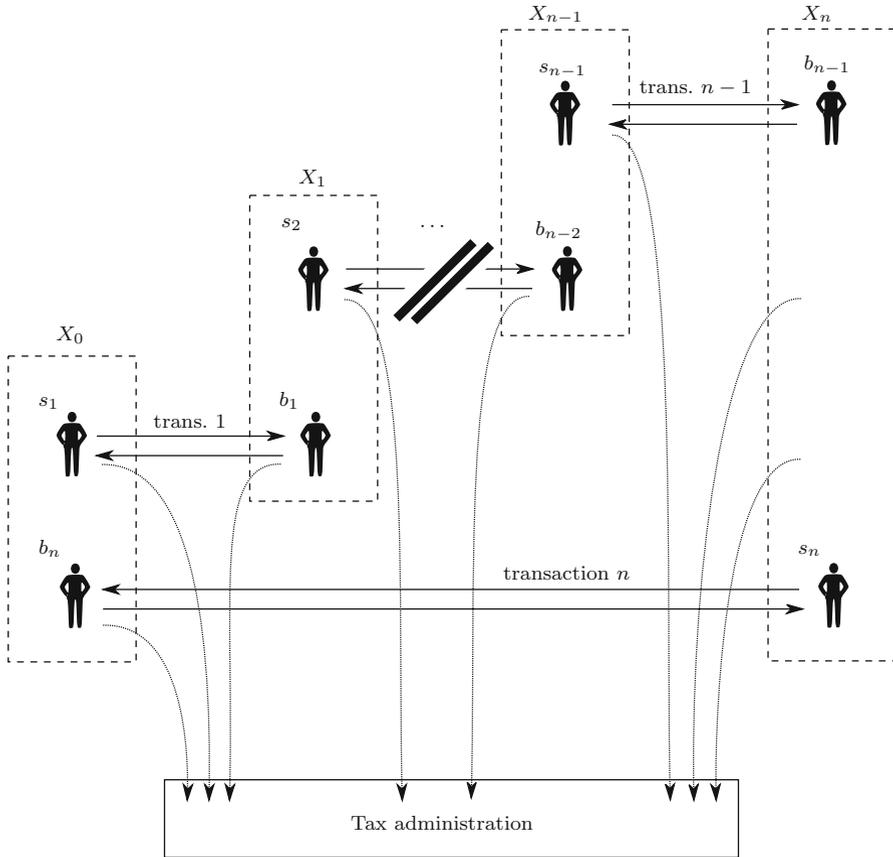
Fig. 2 Swap scheme with n nodes

monitoring and diagnostic processes should be reactive, based on reception of new observations. Therefore, a more plausible monitoring mechanism should look like the following *event-condition-action* (ECA) rule:

(E)   when you receive a declaration,
(C)   if it is suspicious,
(A)   trigger the diagnostic process.

Finally, the diagnostic process should consider the whole family of scenarios that are associated with a symptom, and should consider that there may be *missing information*. One way to proceed in this respect is to integrate a solution similar to EBA, i.e. of generating potential scenarios when needed. Relevant known facts are used to fill fitting scenarios belonging to this family, pruning impossible (according to logic constraints) or implausible (according to prior commitments) ones. Note that this family can be compiled *offline*, as much as the discriminatory power of the

different factors allow. This information may be used to lead the investigation steps to be acted upon in real-time.

In this scenario, the procedural aspect was not essential, but in general, it may be. In related works, for instance, we built our models using (extensions of) Petri nets (Sileno et al. 2014a, d). Petri nets can be mapped to logic programming using for instance Event Calculus (Shanahan 1999) or similar techniques; this can be related to *composite event recognition* approaches (Artikis et al. 2015) suggesting the use of intermediate caching techniques to improve the search. Another solution would be to instead maintain the process notation, and compute fitness decomposing the family of scenario in a hierarchy of *single-entry-single-exit* (SESE) components (Munoz-Gama et al. 2014).

5.2.1 Computational complexity

Model-based diagnosis (MBD) is known to be a hard computational problem, namely exponential to the number of components of the diagnosed systems (see e.g. Bylander et al. (1991)). For this reason, diagnostic algorithms traditionally focus on minimal diagnoses, i.e. of minimal cardinality (involving minimal subset of faulty components), an approach that is also known as the *principle of parsimony* (Reiter 1987). This principle is not directly applicable to our framework, as the system components are not agent-players, but agent-roles enacted by agent-players; each component is therefore 'invisible' to the observation, and can be tracked only as a mechanism involving individual elements. In other words, individual agent-players do not provide the right granularity for diagnostic reasoning here: failures are often due to coalitions (e.g. sport combines) that are not observable *per se* but through the overall behaviour of the individuals.

Fortunately, it has been shown that the exponential increase of computational burden may still be reduced using a mixture of decomposition techniques and statistical information. In this work, we have postponed this problem, as we focused on justifying the proposed method providing a working example of an application. We can, however, identify next directions to investigate. As we said in the previous section, the family of scenarios associated with a certain alarming event is known in advance. Therefore, some knowledge compilation techniques may produce important advantages, deriving heuristic knowledge for heuristic problem-solvers, without restarting from first principles (Chandrasekaran and Mittal 1983; Console et al. 1996). Statistical information may instead be used to focus only on a limited set of most probable *leading* hypotheses (de Kleer and Williams 1989). It has been also suggested to control complexity by using hierarchical models, i.e. models with different levels of abstraction (Mozetič 1991; Chittaro and Ranon 2004; Stern et al. 2014). This is in principle directly possible with agent-roles. All these aspects remain to be investigated.

6 Conclusion and further developments

As already stated in the title, this paper is meant to describe an exercise of computational implementation, targeting a specific problem, exploiting part of the conceptual framework presented in previous works (Boer and van Engers 2011a, b). To have a better focus, we have neglected many other practical and theoretical aspects that have been investigated in parallel, and that should be taken into account to get the full picture. For instance, for the *representation* of agent-roles, we have identified fundamental normative components in *positions*, defined towards another party, in the tradition of Hohfeld's analytic framework (Sileno et al. 2014c), and towards the environment, for practical reasoning purposes (Sileno et al. 2015). We have investigated the *acquisition* of agent-roles starting from UML-like diagrams (Sileno et al. 2014a) and from interpretations of narratives (Sileno et al. 2014d). In these works we worked with (extensions of) Petri nets, introduced, amongst others reasons, to create a natural convergence to the usual notation used for business process models.

On the other hand, this simplification allowed appreciation of the problems of settling a real-time model-based diagnosis activity in operations instead. It is easy to imagine further developments from the insights gained from this exercise. We will just name a few of them: a formalization of the *contrast* operation; the 'compilation' of the collected scenarios in knowledge bases optimized for monitoring and for diagnosis; the interface of EBA with backward-chaining, in order to take into account competing scenarios and the possibility of missing information; the possibility of composing multiple scenarios via planning, taking into account diversional behaviours (this would not be possible with diagnostic systems not relying on models); an investigation on the resulting computational complexity.

References

Artikis A, Sergot M, Paliouras G (2015) An event calculus for event recognition. IEEE Trans Knowl Data Eng 27(4):895–908

Boer A, van Engers T (2011) An agent-based legal knowledge acquisition methodology for agile public administration. In: Proceedings of the 13th international conference on artificial intelligence and law (ICAIL 2011). ACM Press, New York, pp 171–180

Boer A, van Engers T (2011b) Diagnosis of multi-agent systems and its application to public administration. In: Business information systems workshop, Springer, Lecture Notes in business information processing, vol 97, pp 258–269

Bylander T, Allemang D, Tanner MC, Josephson JR (1991) The computational complexity of abduction. Artif Intell 49:25–60

Chandrasekaran B, Mittal S (1983) Deep versus compiled knowledge approaches to diagnostic problem-solving. Int J Man-Mach Stud 19:425–436

Chittaro L, Ranon R (2004) Hierarchical model-based diagnosis based on structural abstraction. Artif Intell 155(1–2):147–182

Console L, Dressier O (1999) Model-based diagnosis in the real world: lessons learned and challenges remaining. In: Proceedings of the 16th international joint conference on artificial intelligence (IJCAI 1999) 2:1393–1400

Console L, Torasso P (1991) A spectrum of logical definitions of model-based diagnosis. Comput Intell 7(3):133–141

Console L, Dupré DT, Torasso P (1989) A theory of diagnosis for incomplete causal models. In: Proceedings of the 11th international joint conference on artificial intelligence (IJCAI 1989), pp 1311–1317

Console L, Portinale L, Dupré DT (1996) Using compiled knowledge to guide and focus abductive diagnosis. IEEE Trans Knowl Data Eng 8(5):690–706

Cox PT, Pietrzykowski T (1986) Causes for events: their computation and applications. In: Deductive databases, planning, synthesis—8th international conference on automated deduction, vol LNCS 230, pp 608–621

Dastani M, van Riemsdijk MB, Hulstijn J, Dignum F, Meyer JJC (2005) Enacting and deacting roles in agent programming. In: Proceedings of the 5th international workshop on agent-oriented software engineering (AOSE 2004), vol 3382, pp 189–204

Davis R (1984) Diagnostic reasoning based on structure and behavior. Artif Intell 24:347–410

Governatori G (2013) Business process compliance: an abstract normative framework. IT Inf Technol 55(6):231–238

Heckerman DE, Horvitz E, Middleton B (1993) An approximate nonmyopic computation for value of information. IEEE Trans Pattern Anal Mach Intell 15(3):292–298

Hutchins E (2010) Enaction, imagination, and insight. Enaction: towards a new paradigm in cognitive science. MIT Press, Cambridge, pp 425–450

Jiang J, Aldewereld H, Dignum V, Tan Yh (2014) Compliance checking of organizational interactions. ACM Trans Manag Inf Syst 5(4):1–24

Kafali O, Torroni P (2012) Exception diagnosis in multiagent contract executions. Ann Math Artif Intell 64(1):73–107

Kalech M (2012) Diagnosis of coordination failures: a matrix-based approach. Auton Agents Multi-Agent Syst 24:69–103

de Kleer J, Williams BBC (1987) Diagnosing multiple faults. Artif Intell 32:97–130

de Kleer J, Williams BC (1989) Diagnosis with behavioral modes. In: Proceedings of the 11th international joint conference on artificial intelligence (IJCAI 1989), pp 1324–1330

Lucas PJ (1998) Analysis of notions of diagnosis. Artif Intell 105(1–2):295–343

Mozetič I (1991) Hierachical model-based diagnosis. Int J Man-Mach Stud 35(3):329–362

Munoz-Gama J, Carmona J, Van Der Aalst WMP (2014) Single-entry single-exit decomposed conformance checking. Inf Syst 46:102–122

Pipattanasomporn M, Feroze H, Rahman S (2009) Multi-agent systems in a distributed smart grid: design and implementation. In: 2009 IEEE/PES power systems conference and exposition, pp 1–8

Preist C, Eshghi K, Bertolino B (1994) Consistency-based and abductive diagnoses as generalised stable models. Ann Math Artif Intell 11:51–74

Reiter R (1987) A theory of diagnosis from first principles. Artif Intell 32(1):57–95

Roos N, ten Teije A, Witteveen C (2003) A protocol for multi-agent diagnosis with spatially distributed knowledge. In: Proceedings of the 2nd international conference on autonomous agents and multiagent systems (AAMAS 2003), p 655

Sadri F (2012) Logic-based approaches to intention recognition. In: Chong NY, Mastrogiovanni F (eds) Handbook of research on ambient intelligence and smart environments: trends and perspectives. IGI Global, Hershey, pp 346–375

Shanahan M (1999) The event calculus explained. Artif Intell Today LNAI 1600:409–430

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423

Sileno G, Boer A, van Engers T (2014a) From inter-agent to intra-agent representations: mapping social scenarios to agent-role descriptions. In: Proceedings of the 6th international conference on agents and artificial intelligence (ICAART 2014), SciTePress Science and Technology Publications, pp 622–631

Sileno G, Boer A, van Engers T (2014b) Implementing explanation-based argumentation using answer set programming. In: Proceedings of the 11th international workshop on argumentation in multi-agent systems (ArgMAS 2014)

Sileno G, Boer A, van Engers T (2014c) On the interactional meaning of fundamental legal concepts. In: Hoekstra R (ed). Legal knowledge and information systems. JURIX 2014: the twenty-seventh annual conference, IOS Press, Amsterdam, vol FAIA 271, pp 39–48

Sileno G, Boer A, Van Engers T (2014d) Legal knowledge conveyed by narratives: towards a representational model. In: Proceedings of the workshop on computational models of narrative (CMN 2014), pp 182–191

Sileno G, Boer A, van Engers T (2015) Commitments, expectations, affordances and susceptibilities: towards positional agent programming. In: Proceedings of the 18th conferences on principles and practice of multi-agent systems (PRIMA 2015), vol LNCS 9387, pp 687–696

Stern R, Kalech M, Elimelech O (2014) Hierarchical diagnosis in strong fault models. In: Proceedings of the 25th international workshop on principles of diagnosis (DX 2014)