Use and acquisition of causal relations for decision making under uncertainty using imperfect information games

by

Mauricio Gonzalez Soto

PhD Thesis Proposal

Computer Science Department

Instituto Nacional de Astrofísica, Óptica y Electrónica
Tonantzintla, Puebla, Mexico
January 17th 2018

Supervisors:

Dr. Hugo Jair Escalante Balderas
INAOE, Mexico

Contents

1 Introduction **1**
1.1 Outline . . . 2

2 Main Concepts **4**
2.1 Attempts to define Causality . . . 4
2.2 Definition of Causality . . . 4
2.3 Representation into a Directed Acyclic Graph . . . 5
2.3.1 Relation between the graph and probabilities . . . 5
2.4 Causal Graphical Models . . . 6
2.4.1 The identifiability problem . . . 6
2.4.2 Do-Calculus . . . 6
2.4.3 Use of Causal Graphical Models over other causal models . . . 8
2.5 Decision Theory . . . 8
2.5.1 Elements of a Decision Problem . . . 8
2.5.2 Axioms of Coherence . . . 9
2.5.3 Subjective Probability and Utility . . . 11
2.5.4 Decision Criteria for Decision Problems . . . 12
2.6 Game Theory . . . 12
2.6.1 Building blocks . . . 13

3 Problem Statement **16**
3.1 Problem Statement . . . 16
3.2 Motivation . . . 16
3.3 Justification . . . 16
3.4 Research Question . . . 17
3.5 Hypothesis . . . 18
3.6 Objectives . . . 18
3.6.1 General Objectives . . . 18
3.6.2 Specific Objectives . . . 18
3.6.3 Limitations . . . 18
3.6.4 Extensions . . . 19

4 Related Work **20**

5 Methodology **23**
5.1 A guiding principle for a Causal Decision Problem . . . 23
5.2 Modelling a Decision Problem under Uncertainty as a Game . . . 23
5.2.1 Elements of the Game . . . 24
5.3 Belief Formation . . . 24
5.4 Belief updating . . . 25
5.5 Solving first simpler cases . . . 25

5.5.1 Fully knowing the causal graphical model . . . 26
5.5.2 Knowing only the graph structure . . . 26
5.5.3 General case: Model unknown . . . 26
5.6 Balancing exploitationd and exploration . . . 27
5.7 Validation of the proposed methods . . . 27
5.8 Experimental methodology . . . 27
5.8.1 Internal Validation: Reproducibility . . . 28
5.8.2 External Validation: Different problems . . . 28
5.8.3 Concurrent Validation . . . 28
5.9 Working plan . . . 28
5.9.1 Publications plan . . . 29

6 Preliminary Results **30**
6.1 Test scenario . . . 30
6.2 Case 1: The causal model is completely known . . . 31
6.3 Case 2: Only the structure is known . . . 32
6.3.1 Implementation . . . 33
6.4 Case 3 and future work: The model is not known . . . 34

Abstract

We consider decision problems under uncertainty where the options available to a decision maker and the resulting outcome are related through a causal mechanism which is unknown to the decision maker, although he is aware of the causal nature of his environment. We study how a decision maker can learn about this causal mechanism through sequential decision making as well as using current causal knowledge inside each round in order to make better choices had he not considered causal knowledge. We propose a decision making procedure in which an agent holds *beliefs* about her environment which are used to make a choice and then are updated using the observed outcome. As proof of concept, we present an implementation of this causal decision making model and apply it to a simple problem. We show that the model achieves a performance similar to the classic Q-learning while it also acquires a causal model of the environment.

1 Introduction

A fundamental part of intelligent reasoning is being able to make decisions under uncertain conditions (Danks (2014), Lake et al. (2017), Pearl (2018b)). Uncertain conditions as when that a set of choices is made available to a decision maker he can not predict the precise consequence of his decision. For this reason, the decision maker must take into account the stochastic relation between actions[1] and consequences. From the von Neumann-Morgenstern decision making theory we know that if a rational[2] decision maker knows this stochastic relation, then he must*If rationality is assumed* choose the action that maximizes the *expected utility* of a *utility function* whose existence is guaranteed (Von Neumann and Morgenstern (1944)). On the other hand, if a rational decision maker does not know the probabilities of the ocurrence of certain outcomes, or consequences, given his actions, then he must encode his personal, subjective, quantifications of any uncertainties in a probability distribution p and choose the action that maximizes expected utility with respect to this probability distribution(Savage (1954), Bernardo (2000), Gilboa (2009)). If the current *knowledge* of a decision maker is not enough to make good choices, then the decision maker can *learn* from the environment by interacting with it in order to update his quantifications of uncertainty (Bernardo (2000), Peterson (2017)).

Learning by interaction has been extensively studied by computer scientists using the Reinforcement Learning (RL) setting (Sutton and Barto (1998)), but the most common used techniques in this field are purely associative and do not consider any high-level structure of the environment beyond what is expressable in a Markov Decision Process (Garnelo et al. (2016)). A particular case of a *higher level structure*; i.e., beyond associative patterns, is the case of *causal* structure. A causal structure encondes a series of *cause-effect* relations between events and knowing such relations allows a decision maker to add extra knowledge into the uncertainty of his environment and also allows to plan ahead his actions since he can predict what a certain action will cause (Spirtes et al. (2000), Pearl (2018a)).

[1] Actions, options and choices to be used equivalently.

[2] To be explained in full detail later.

Causal reasoning is found at the very core of human reasoning, since it allows us to manipulate our environment and being able to predict effects of a given action (Spirtes et al. (2000)). Causal claims also support reasoning of the form *if...hadn't occurred then...wouldn't have happened*. It is argued by Pearl and Mackenzie (2018) that imagining alternative scenarios is fundamental for humans to make causal reasoning. Since human beings are known to learn causal models in sequential decision making processes (Sloman and Hagmayer (2006), Nichols and Danks (2007), Meder et al. (2010), Hagmayer and Meder (2013), Danks (2014)), and even though this learning is not perfect (Rottman and Hastie (2014)), our hypothesis is that an autonomous agent can learn and use causal information while interacting with an uncertain environment which is governed by a fixed *causal mechanism* which is unknown to the agent.

The proposed way for an agent to learn from repeated interactions is by giving him *beliefs* about the structure of the environment, which are to be used as a *true* causal model, and then are to be updated them after an outcome has been observed. While the standard setting in RL is to model the agent-environment interaction as an agent that moves from one *state* to another inside a model of the environment and observing a reward as these transitions occur, we propose to model it as a *game* between the decision maker and a player called *Nature* which will select his actions from the causal model in response to what the decision maker has chosen. The agent, besides learning a policy to choose actions will also learn a causal model from the environment since the causal model she forms will approximate the true model.

Learning a causal model of the environment allows to extract high-level insights of a phenomena beyond associative descriptions of what is observed. A causal model is able to *explain* why a particular decision was made since it allows to extract the causes and effects of an agent's actions. Once a causal model is acquired, an external user is able to reason about *what...if...* statements that associative methods can not answer. When a decision maker chooses an action out of many, a causal model allows to ask what would've happened if another action was taken without actually performing the alternative action. It is important to be careful when using the word *explanation* since it allows for several uses in language, here we understand *explanation* as used by Woodward (2005) who considers explanations where what is explained depends on other factors via some relationship that holds as a matter of empirical fact rather than *logical* reasons.

Learning a causal model of a certain environment would allow transferability of knowledge into similar domains because the underlying cause-effect structure has been captured and could later be used.

1.1 Outline

In the Section 2 the main concepts that serve as a foundation for this work will be described; in particular, we will describe in detail what causality means. Since this research proposal asumes that a causal model governs an environment, a precise definition of causality will be given along with the mathematical machinery to perform causal inference. Also, since the entity that will be facing causal environments is a *rational* decision maker we will precisely define, axiomatically, what does rationality means and which are the main theories of rational

decision making. At the end of the section, we will describe a mathematical theory which is suitable to modelling interaction between rational decision makers since interactive learning is considered.

After the main concepts have been exposed, in Section 3 a more precise statement of the problem that we propose to attack will be given, along with the Hypothesis that we attempt to defend and believe that is the answer to the problem.

Once the main problem has been stated, we proceed in Section 4 to show previous attempts over this and similar problems.

Finally, we state the proposed solution for the problem as well as showing results of experiments performed which implement a small toy example.

2 Main Concepts

2.1 Attempts to define Causality

Causality has been a difficult concept to define and several attempts have been made. In fact, in mainstream statistics the term *cause* have been nearly banned as documented by Pearl and Mackenzie (2018).

Attempts to understand the meaning of *cause* has troubled philosophers and scientists, from David Hume and John Stuart Mill to modern scientists as Suppes (1970), Cartwright (1983), Spirtes et al. (2000), Pearl (2009), Spohn (2012), Hitchcock (2018). In particular, Suppes and Reichenbach tried to define causation as an increase in the probability of occurrence of one event when another event happened, which is stated as X causes Y if $P(Y|X) > P(Y)$, but this definition was flawed because the reason behind such increase could happen for several other reasons, such as Y being a cause of X, or both having a common cause Z (Pearl and Mackenzie (2018)).

Several definitions of Causality exist in diverse contexts with diverse applications in mind, such as the Topological Causality for Dynamical Systems (Harnack et al. (2017)), Lamport's Causality (Lamport (1978)), Granger's Causality for Time Series (Granger (1969)), Suppes' Causality (Suppes (1970)) each with its own mathematical framework. A review of several of this theories, including Suppes' and Granger's can be found in Holland et al. (1985).

In some works on Causality (Spirtes et al. (2000), Pearl (2009)) the term *causality* is defined in terms of common language and then the mathematical machinery required to model causality is provided. In this way, causality becomes defined as what is modeled by causal models, which are themselves defined as to model causality. This apparent circularity is deeply studied by Woodward (2005). In fact, Judea Pearl never gives a formal definition of causality, rather he prefers to build the machinery required to answer *causal queries* (Pearl (2009), Pearl (2018a)). Causality, as probability, can be thought of either as an objective phenomena that *exists* in the environment, or as a mental construct that enables us to handle complex situations.

Since causal claims in common language can have different logical form, in this work we give a *working* definition of causality that allows a formal language for stating and manipulating causal concepts and we will take the given definition as axiomatic.

2.2 Definition of Causality

Causality is understood as a *stochastic* relation between *events*. Some event (or events) *causes* another event to occurr. The formal definition of Causality that will be adopted in this work is the definition provided in Spirtes et al. (2000).

Definition 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ a finite probability space, and consider a binary relation $\rightarrow \subseteq \mathcal{F} \times \mathcal{F}$ which is:

- Transitive: If $A \rightarrow B$ and $B \rightarrow C$ for any $A, B, C \in \mathcal{F}$ then $A \rightarrow C$.
- Irreflexive: For all $A \in \mathcal{F}$ it doesn't hold that $A \rightarrow A$.

- Antisymmetric: For $A, B \in \mathcal{F}$ such that $A \neq B$ if $A \to B$ then it doesn't hold that $B \to A$.

We will say that A is *a cause* of B (or that A causes B, A is the cause and B is the effect) if $A \to B$. It is important to note that an event may have more than one cause, and that not necesarily each one of this causes is sufficient to produce the effect.

We will assume that for any event $A \in \mathcal{F}$ there are no causes lying outside $\mathcal{F}$. It will also be assumed that the relations expressed by $\to$ are the only causal relations in the environment. This assumption can be interpreted as not allowing *intermediate* events between events known to be causally related. For example, if in our space we have that striking a match causes fire, we will not allow into consideration the underlying chemical reactions that cause fire from friction. In this sense, we will say that striking a match is a *direct cause* of fire and this will be the only causes considered herein (Spirtes et al. (2000)).

2.3 Representation into a Directed Acyclic Graph

The causal relations contained in $\to$ can be summarized in a graph $G = (V, E)$ in the following way: If $A \to B$ then the graph must contain a node $A \in V$ representing A, a node $B \in V$ representing B and a directed edge $e \in E$ connecting the respective nodes in the direction of the causal relation.

Proposition 2.1. Given a causal relation $\to$ as in Definition 2.1 then the graph that is obtained by considering nodes for events and edges for the causal relations as previously described is a Directed Acyclic Graph.

Proof. The graph is directed since the definition of causality imposes a direction between events; namely, a direction between the cause and the effect. To see that the graph is acyclic, suppose that a cycle $A \to B \to C \to A$ exists, this would imply, because of transitivity, that $A \to A$, which can not be since the relation is irreflexive. □

Notice that since the graph is finite, there exist some nodes that does not have causes, which are called *exogenous*. If an A event is caused by some other event, then we say it is *endogenous* and we denote the set of its causes as $Pa(A)$. It is proven in **?** that at least one exogenous node exists in a causal graph.

2.3.1 Relation between the graph and probabilities

Given a Directed Acyclic Graph it is possible to obtain a probability measure that expresses the conditional independence relations that are expressed in the graph (Koller and Friedman (2009)), which we will call $P_{\mathcal{G}}$. For the DAG built from the causal relations, we require that its correspondent $P_{\mathcal{G}}$ satisfies the following conditions (Spirtes et al. (2000)):

- Markov Causality: an event V (or node in $\mathcal{G}$) is independent of every other event A such that A isn't either a cause nor an effect of V given the causes of V.

- Causal Minimality: No proper sub-graph of $\mathcal{G}$ satisfies the Markov Causality condition.

- Causal Faithfulness: The Markov Causal condition contains all of the conditional independence statements expressed by the DAG $\mathcal{G}$.

Also, an extra pair of conditions, the first one known as Causal Sufficiency, which is about the nature of the model: for any variable X in the model $\mathcal{G}$ there are no causes of X outside of the model $\mathcal{G}$ (Spirtes et al. (2000), Pearl (2009), Sucar (2015)). The second required condition is that the causal mechanism remains unchanged after interventions, this is called in (Woodward (2005)) as *invariantness*, this basically means that an interventor does not *breaks* the causal mechanism when intervening upon it. This conditions are required in an axiomatic fashion so they will be taken as they are without questioning.

2.4 Causal Graphical Models

A causal graphical model (CGM) consists of a set of random variables $\mathcal{X} = \{X_1, ..., X_n\}$, an acyclic directed graph (DAG) whose nodes are in correspondence with the variables in $\mathcal{X}$ and whose edges represent relations of cause-effect. Also, the model, which up to this point is nothing more than a Bayesian Network with causal semantics, is enriched with an operator named $do()$ which is defined over graphs and whose action is described as follows: given $\mathbf{X} \subseteq \mathcal{X}$ and $\mathbf{x} = \{x_{i_1}, x_{i_2}, ..., x_{i_j}\} \in Val(\mathcal{X})$ the action $do(\mathbf{X} = \mathbf{x})$ corresponds to assign to each $X_j \in \mathbf{X}$ the value $x_{x_{i_j}}$ and to delete every incoming edges into the node corresponding to each X_j in the graph $\mathcal{G}$ (Koller and Friedman (2009), Sucar (2015)) To apply the $do()$ operator over a variable (or set of variables) is also called as an *intervention* over the variable. It is this interventional operator which separates associative models from causal models, since it provides the solution of the "probability raising" attempts to define causality, since if X is a cause for Y then $P(Y|do(X)) > P(Y)$ (Pearl and Mackenzie (2018)).

It is required that the probability distribution that results from an intervention over a variable is Markov compatible with the graph; this is, the resulting interventional distribution is equivalent to the product of the conditional probability of every variable given its parents in the intervened graph (Sucar (2015)).

2.4.1 The identifiability problem

Under what conditions can causal inquiries be answered in terms of purely *observational* data? It is known that if the Markov Causal condition and Causal Faithfulness hold, then it is possible to identify the causal graph up to related variables without direction (Markov equivalence) (Peters et al. (2012), Mooij et al. (2016)).

2.4.2 Do-Calculus

The Do-Calculus (Pearl (1995), Pearl (2009)) is a set of rules for manipulating probabilistic statements that involve interventions and, under certain conditions, allow them to be transformed into statements that do not involve interventional data.

Some notation is important to be mentioned: Consider a causal graphical model $\mathcal{G}$ and X, Y, Z disjoint sets of nodes of $\mathcal{G}$. We denote by $\mathcal{G}_{\bar{X}}$ the graph that is obtained by deleting from $\mathcal{G}$ all of the edges that enter nodes in X. In the same way, $\mathcal{G}_{\underline{X}}$ is the graph obtained by

deleting the edges that emerge from X. Finally, $\mathcal{G}_{\underline{Z}\bar{X}}$ is the graph obtained by deleting edges incoming into X and outgoing from Z.

Theorem 2.1. (Pearl (2009))
Let $\mathcal{G}$ a CGM and $P_{\mathcal{G}}$ the probability measure induced by the model; then, for disjoint sets of nodes X, Y, Z, W it holds:

- If for the graph $\mathcal{G}_{\bar{X}}$ it holds that Y is conditionally independent from Z given X and W, then

$$P_{\mathcal{G}}(Y = y|do(X = x), Z = z, W = w) = P_{\mathcal{G}}(Y = y|do(X = x), W = w).$$

- If for the graph $\mathcal{G}_{\bar{X}\underline{Z}}$ it holds that Y is conditionally independent from Z given X and W, then

$$P(Y = y|do(X = x), do(Z = z), W = w) = P(Y = y|do(X = x), Z = z, W = w).$$

- Let $Z(W)$ the set of nodes in Z that aren't ancestors of any node in W in the graph $\mathcal{G}_{\bar{X}}$. If Y is conditionally independent from Z given X and W in the graph $G_{\bar{X},Z(\bar{W})}$, then

$$P(Y = y|do(X = x), do(Z = z), W = w) = P(Y = y|do(X = x), W = w).$$

Proof. See Pearl (2009). □

Theorem 2.2. Peters et al. (2017)
The following statements hold:

- The Do-Calculus is complete; this is, sufficient for deriving every identifiable interventional distribution (Huang and Valtorta (2006), Shpitser and Pearl (2006)).
- There exists an algorithm capable of finding all of the identifiable interventions (Tian (2002), Huang and Valtorta (2006)).
- A necessary and sufficient criteria exists for the identifiability of interventional distributions (Shpitser and Pearl (2006), Huang and Valtorta (2006)).

Proof. See Peters et al. (2017). □

The Do-Calculus rules provide a solution for the identifiability problem:

Corollary 2.1. (Pearl (2009)).
A distribution $q = P(y_1, ..., y_k|do(x_1), ..., do(x_n))$ is identifiable in a causal graphical model $\mathcal{G}$ if there exists a finite sequence of transformations, where each one of them corresponds to any of the rules of the Do-Calculus, that reduce q to a purely observational expression.

2.4.3 Use of Causal Graphical Models over other causal models

We choose to use specifically a Causal Graphical model because of the mathematical machinery that exists around them and because Graphs are a model of Cognitive Processes, such as causal reasoning (Glymour (2003), Danks (2014)). In this way, if an agent *knows* that a certain environment is governed by causal relations, then a causal graphical model can be used to model the causal reasoning of the agent.

2.5 Decision Theory

A Decision Problem under Uncertainty is a situation in which an agent must choose one out of many actions with uncertain consequences. The most common theories for Decision Making under Uncertainty are those from Von Neumann and Morgenstern (1944) and Savage (1954). In the former theory it is assumed that the decision maker knows the stochastic relation between actions and consequences [3], and in that case the theory guarantees that the decision maker behaves as if he maximizes the expected value of an utility function. On the other hand, if the decision maker doesn't know the probabilities of observing an outcome given a chosen action, then Savage's theory guarantees that the decision maker behaves as if he had in mind a *subjective*[4] probability distribution, a utility function and chooses the action which maximizes the expected utility with respect to that subjective probability distribution and utility function.

Both von Neumann and Savage's Decision Under Uncertainty theories are known as the *classic* theories in other areas, mostly in Economics. It is important to say that other Decision Making theories exist, such as Prospect Theory (Kahneman and Tversky (1979)), Case-Based Decision Theory (Gilboa and Schmeidler (1995)) among others that are out of the scope of this work.

For what remains of this section, we follow Bernardo (2000) in his exposition of Savage's theory since it is the most suitable for our purposes because we will be considering decision makers that do not know all of their environment. Savage's Decision Making theory, together with von Neumann's theory is what it is commonly known as *classical* decision making.

2.5.1 Elements of a Decision Problem

A Decision Problem under Uncertainty is composed by:

- A set Ω of *states*.
- A set $\mathcal{A} = \{a_i : i \in I\}$ of actions.
- For each action a_i, a partition $E_i = \{E_j : j \in J\}$ of Ω. Define $\mathcal{E} = \cup_{i \in I} E_i$.
- For each action a_i, a set of consequences $C_i = \{c_j : j \in J\}$. Define $\mathcal{C} = \cup_{i \in I} C_i$

[3] Also known as decision under risk (Binmore (2008), Peterson (2017)).

[4] Subjective here is a technical term meaning that the function is internal for the agent.

- A preference relation $\succeq$ defined over $\mathcal{A}$ which represent the decision-maker whishes. It is assumed that the decision maker will choose among his options according to his preferences, which are represented by $\succeq$.

Since we are considering that a decision maker chooses an action $a \in \mathcal{A}$, and then an uncertain event E will occur which will then produce the consequence c, we will identify actions in $\mathcal{A}$ as $a_i = \{c_j | E_j : j \in J\}$. As a technical assumption required for all of the math to work, we require that elements of $\mathcal{C}$ belong to $\mathcal{A}$. This is achieved by adding into A elements of the form $\{c|\Omega\}$ for $c \in \mathcal{C}$. This being said, we can extend the preference relation $\succeq$ to elements of $\mathcal{C}$ and therefore indistinctively write $c_1 \succeq c_2$ for $c_1, c_2 \in \mathcal{C}$ or $a_1 \succeq a_2$ for $a_1, a_2 \in \mathcal{A}$. We require that $\mathcal{E}$ is an *algebra*[5].

Definition 2.2. From the prerefence relation $\succeq$ we can derive some other relations between actions:

(i) $a_1 \sim a_2 \Leftrightarrow a_1 \succeq a_2$ and $a_2 \succeq a_1$

(ii) $a_1 \succ a_2 \Leftrightarrow a_1 \succeq a_2$ and it does not hold that $a_2 \succeq a_1$.

Definition 2.3. A relation between events can be derived from $\succeq$ in the following way:

$$E \succeq F \Leftrightarrow \text{ for all } c_1 \succeq c_2 \text{ it holds that } \{c_2|E, c_1|E^c\} \succeq \{c_2|F, c_1|F^c\}.$$

In this case, we say that F is **not more likely** than E. It can easily be shown that $\Omega \succ \emptyset$, and a relation $\sim$ for events is defined in an analog way than for consequences.

Definition 2.4. Given some $G \succ \emptyset$ we define a conditional preference relation $\succeq_G$ as follows:

(i) $a_1 \succeq_G a_2 \Leftrightarrow$ for all a $\{a_1|G, a|G^c\} \succeq \{a_2|G, a|G^c\}$.

(ii) $E \succeq_G F \Leftrightarrow$ for $c_1 \succeq_G c_2 \{c_2|E, c_1|E^c\} \succeq_G \{c_2|F, c_1|F^c\}$

Definition 2.5. Two events $E, F \in \mathcal{E}$ are said to be independent if and only if for all $c, c_1, c_2 \in \mathcal{C}$ it holds:

(i) $c \succ \{c_2|E, c_1|E^c\} \Rightarrow c \succ_F \{c_2|E, c_1|E^c\}$.

(ii) $c \succ \{c_2|F, c_1|F^c\} \Rightarrow c \succ_E \{c_2|F, c_1|F^c\}$

2.5.2 Axioms of Coherence

We require the following axioms for the preference relation $\succeq$ which we will call Axioms of Coherence or Rationality, and define a Rational decision maker as a decision maker who will choose according to a preference relation and whose preferences satisfy the axioms (Bernardo (2000), Gilboa (2009)).

Axiom 1: Comparability.

[5]An algebra of sets over X is a family of sets $\mathcal{F} \subseteq \mathcal{P}(X)$ such that $X \in \mathcal{F}$, $A^c \in \mathcal{F}$ for any $A \in \mathcal{F}$ and for $A, B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$ (Ash and Doleans-Dade (2000))

(i) There exists consequences c_1, c_2 such that $c_1 \succ c_2$.

(ii) For all consequences c_1, c_2 and events $E, F \in \mathcal{E}$ either $\{c_2|E, c_1|E^c\} \succeq \{c_2|F, c_1|F^c\}$ or $\{c_2|F, c_1|F^c\} \succeq \{c_2|E, c_1|E^c\}$

We require that the decision maker is able to strictly prefer at least one action over another, because if this weren't the case then there would not be a decision problem, since any action would lead to the same consequence. It is also required that the decision maker has preferences over simple options.

Axiom 2: Transitivity.

(i) $a \succeq a$ for all $a \in \mathcal{A}$.

(ii) If $a \succeq b$ and $b \succeq c$ for $a, b, c \in \mathcal{A}$, then $a \succeq c$.

Even though transitivity is a very intuitive requirement, its necessity is better explained in terms of *money pumps*: consider c_1, c_2 and c_3 such that $c_1 \succ c_2$ and $c_2 \succ c_3$ but $c_3 \succ c_1$. In this case, the decision maker would be willing to pay a strictly positive ammount to have c_3 over c_1, and to pay $y > 0$ to have c_2 rather than c_3, and an amount $z > 0$ in order to have c_1 rather than c_2, and then again to pay to have c_3 over c_1...

Axiom 3: Consistency.

(i) If $a_1 \succeq a_2$ then for all $G \succ \emptyset$ $a_1 \succeq_G a_2$.

(ii) If for some $c_1 \succ c_2$ it holds that $\{c_2|E, c_1|E^c\} \succeq \{c_2|F, c_1|F^c\}$ then $E \succeq F$.

(iii) If for some $c \in \mathcal{C}$ and $G \succ \emptyset$ it holds that $\{c_1|G, c|G^c\} \succeq \{c_2|G, c|G^c\}$ then $c_1 \succeq_G c_2$.

The intuition behind condition (i) is that preferences between *pure* consequences should not be affected by information about the uncertain events in $\mathcal{E}$. Conditions (ii) and (iii) make definitions 2.3 and 2.4 well defined concepts. In particular, condition (ii) formalises the idea that preferences should not depend on relative likelihood. Condition (iii) commonly known as the *sure thing principle* (Bernardo (2000)).

Axiom 4: Existence of Standard Events.

There exists a sub-algebra $\mathcal{S}$ of $\mathcal{E}$ and a function $\mu : \mathcal{S} \to [0, 1]$ such that:

(i) $S_1 \succeq S_2$ if and only if $\mu(S_1) \geq \mu(S_2)$.

(ii) If $S_1 \cap S_2 =$ then $\mu(S_1 \cup S_2) = \mu(S_1) + \mu(S_2)$.

(iii) For any number $\alpha \in [0, 1]$ and independent events E, F, there is an event S, which is called a *standard event*, such that $\mu(S) = \alpha$ and S is independent from E and from F.

(iv) iv) If S_1 is independent from S_2 then $\mu(S_1 \cap S_2) = \mu(S_1)\mu(S_2)$.

(v) v) If E independent of S, S independent from F and E independent from F then $E \sim S$ implies that $E \sim_F S$.

This axioms means that the decision maker has a *mental procedure* to generate a probability distribution where one event is more likely than other if and only if it has a higher probability. From the previous axioms, the preferences and uncertainties of a decision maker can be quantified up to small intervals, but in order to assign a unique value the following axiom is required.

Axiom 5: Precise measurement.

(i) If $c_1 \succeq c \succeq c_2$ then there exists a standard event S such that $c \sim \{c_2|S, c_1|S^c\}$.

(ii) For each event E there exists a standard event S such that $E \sim S$.

For a richer discussion, critiques and extensions of the coherence axioms we refer the reader to Bernardo (2000), Binmore (2008), Gilboa (2009), Wakker (2010), Peterson (2017).

2.5.3 Subjective Probability and Utility

Subjective probability is a well defined mathematical object that, contrary to the frequentist approach to probability, represents the degree of belief an agent has over the occurrence of some phenomena. Frequentist probability can not answer questions like *what is the probability of a war in the middle east?* since it is not a phenomena that can be repeated infinite times in an independent way (Gilboa (2009), Peterson (2017)).

Subjective approachs to probability were developed by Bruno De Finetti (De Finetti (1930)) and Frank Ramsey in 1926, however a more precise formulation was given by Savage in his famous book "The Foundations of Statistics" (Savage (1954)). The formulation shown here is the modern version of Savage's Theory as exposed by Bernardo (2000).

Definition 2.6. Given a preference relation $\succeq$, we define the (subjective) **probability** of an event $E \in \mathcal{E}$ as the real number $\mu(S)$ associated with the standard event S such that $E \sim S$.

The subjective probability thus defined satisfies all of the Kolmogorov's axioms of probability (Bernardo (2000)). So, we can use all of the mathematical machinery for classic probability measures, including the definition of conditional probability, Bayes' theorem, etc.

Not every belief about uncertainties can be represented by a probabiity measure, a subjective degree of belief for an agent must be rational as will be shown soon.

As a technical requirement we need to consider two extra options, one that is preferred over every other consequence, which will be denoted as c^* and other consequence c_* which the agent prefers every other consequence before c_*. This consequences need not have an interpretation, but they can be thought of as *heaven* and *hell* respectively for obvious (at lest in the Western hemisphery) reasons. Critiques of these extra technical consequences can be found in Binmore (2008), Peterson (2017).

Definition 2.7. Consequences c^* and c_* are called respectively best and worst (extreme consequences) if for any other consequence $c \in \mathcal{C}$ then $c^* \succeq c \succeq c_*$. Decision problems in which we add extreme consequences are called bounded decision problems.

Definition 2.8. Given a preference relation $\succeq$ in a bounded decision problem we define the canonical **utility** $u(c) = u(c|c_*, c^*)$ of a consequence $c \in \mathcal{C}$ relative to the extreme consequences c_*, c^*, as the real number $\mu(S)$ associated with any standard event S such that $c \sim \{c^*|S, c_*|S^c\}$.

2.5.4 Decision Criteria for Decision Problems

Proposition 2.2. For any bounded decision problem with extreme consequences c_* and c^* it holds:

(i) For all $c \in \mathcal{C}$, $u(c|c_*, c^*)$ exists and is unique.

(ii) The value of $u(c|c_*, c^*)$ is unafected by the occurrence of any event $G \succ \emptyset$.

(iii) $0 = u(c_*|c_*, c^*) \leq u(c|c_*, c^*) \leq u(c^*|c_*, c^*) = 1$.

Definition 2.9. For a bounded decision problem and any event $G \succ \emptyset$ and $a = \{c_j|E_j : j \in J\}$ we define the conditional expected utility of a as

$$\bar{u}(a|c_*, c^*, G) = \sum_{j \in J} u(c_j|c_*, c^*)P(E_j|G).$$

If $G = \Omega$ we simply denote $\bar{u}(a|c_*, c^*, G)$ as $\bar{u}(a|c_*, c^*)$

Theorem 2.3. For any bounded decision problem with extreme consequences $c^* \succ c_*$ and any $G \succ \emptyset$

$$a_1 \succeq_G a_2 \Leftrightarrow \bar{u}(a_1) \geq \bar{u}(a_2).$$

Proof. See Bernardo (2000) Chapter 2 "Foundations", Proposition 2.22. □

This result means that for any decision maker whose preferences satisfies axioms 1 through 5 the only election criteria that is compatible with the axioms is the maximization of expected utility and thus establishes a *normative* criteria for decision making for rational agents. This result establishes a complete ordering of the options, but it does not guarantees the existence of an option for which the expected utility is maximum and further mathematical assumptions are required over the utility function u in order to guarantee the existence of the maxmimum utility option (Bernardo (2000)).

Notice that the expectations are taken with respect to the subjective probability measure, and in fact Savage's general result as stated in Gilboa (2009) guarantees the existence of a utility function and a probability measure such that the preference relation is represented by the expected utility of that function with respect to that probability measure. Here we defined the subjective probability directly from the coherence axioms.

In the particular case where the algebra $\mathcal{E}$ contains only the set Ω as its only element we say that it is a decision problem without uncertainty and in this case applying the previous Theorem we obtain the decision criteria which consist in maximizing the utility function.

The result extends previous work of Von Neumann and Morgenstern (1944) where it is assumed that the decision maker knows the probabilties of the uncertain events.

2.6 Game Theory

A *game* arises when two or more *rational* decision makers, or players, have to make decisions in a situation in which the outcome for each player is partly determined by the choices of other

players (Binmore (2008)). Game Theory is an area of Mathematics which is used to model games. The basic assumptions of the theory is that each decision maker is rational, and they take into account their knowledge or expectations of other decision makers behavior.

In this section we will review the basic models of Game Theory that are required to this thesis proposal, which are the normal-form game, and the extensive-form game.

2.6.1 Building blocks

The normal-form game models a situation in which two or more agents interact and where each one of them will choose an action simultaneously. Also, it is assumed that all of the relevant aspects of the game are known for each player.

We follow the exposition from Osborne and Rubinstein (1994), Binmore (2007), Shoham and Leyton-Brown (2008).

Definition 2.10. A normal-form game is a tuple $(N, A, (\succeq_i)_{i \in N})$ where:

- N is a finite set of players.
- $A = A_1 \times ... \times A_n$ where each A_i is the set of available actions for player i. Each $a = (a_1, ..., a_n) \in A$ is called an action profile.
- For each player $i \in N$, a preference relation $\succeq_i$ defined over A.

A *strategy* or action profile is an element $a \in A$.

Because of the decision-making criteria shown in the last section we know that each player can replace his preferences $\succeq_i$ for a utility function u_i.

The prisoner's dilemma is the most common example of a normal-form game, where two suspects must either confess or remain silent about a crime and they can't communicate with each other. In this game, both of the players know the consequences of each *outcome* and the rewards (utilities) associated with each outcome.

In some cases, the players do not know all of the relevant aspects of a game, such as the payoffs or the available actions to other players. It is this situation that is modeled by normal-form Bayesian games (Osborne and Rubinstein (1994), Shoham and Leyton-Brown (2008)).

Definition 2.11. A normal-form Bayesian game consists of:

- A set of states $\mathcal{S}$.
- A finite set of players N.
- For each player: a set of actions A_i and as in the previous definition we define $A = A_1 \times \cdots \times A_n$.
- For each player: a finite set T_i of *signals* that are observable to player i.
- For each player: a signal function $\tau_i : \mathcal{S} \to T_i$.
- For each player: a probability measure P_i defined over $\mathcal{S}$ such that $P(\tau_i^{-1}(t_i)) > 0$ for $t_i \in T_i$.

- For each player: a rational preference relation $\succeq_i$ defined over the set of probability measures defined over $A \times \mathcal{S}$.

The interpretation of the elements of this model is as follows: the set $\mathcal{S}$ contains the descriptions of the relevant characteristics for all players, each player has *a priori* beliefs over this characteristics, which are stated by the probability measure P_i. When a play is to be realized, the world is in some $\omega \in \Omega$ and each player observes his $\tau_i(\omega)$. If a player observes the signal $t_i \in T_i$ then he concludes that the real state of the world is in the set $\tau_i^{-1}(t_i)$. Player i updates his beliefs over $\omega \in \Omega$ to $P_i(\omega)/P_i(\tau_i^{-1}(t_i))$ if $\omega \in \Omega$ and cero otherwise. Bayesian games are also called *incomplete information* games in the literature which is different from the *imperfect information* game which will be soon defined.

The normal-form game models situations where players select their actions simultaneously, or when it does not matter who makes an action first. More realistic and complex situations involve some notion of temporal structure, or order among the players' actions. The extensive-form game attempts to include this notion of order. The standard references for both of this models are Osborne and Rubinstein (1994), Shoham and Leyton-Brown (2008), Sorin (2003).

Definition 2.12. A perfect information extensive game consists of:

- A finite set N of players.
- A set A of available action.
- A set H of sequences called histories that satisfies:
  - The epty sequence belongs to H.
  - If $(a_k)_{k=1}^K \in H$ (K can be infinite) and $L < K$ then $(a_k)_{k=1}^L \in H$.
  - If $(a_k)_{k=1}^\infty$ satisfies $(a_k)_{k=1}^L \in H$ for every positive integer L, then $(a_k)_{k=1}^\infty \in H$
- If a history $(a_k)_{k=1}^K \in H$ with K finite and if it doesn't exist a $K+1$ such that$(a_k)_{k=1}^{K+1} \in H$ then $(a_k)_{k=1}^K$ is said to be a terminal history. The set of terminal histories is called Z
- A function $\chi : H \to 2^A$ that assigns a set $\chi(h)$ of possible actions to be taken after a history h has occurred.
- A function ρ that assigns to every non terminal history h a player $\rho(h)$ who will choose an action after history h has occurred.
- A function $\sigma : H \times A \to H \cup Z$ that maps history-action pairs into a new history such that if $\sigma(h_1, a_1) = \sigma(h_2, a_2)$ then $h_1 = h_2$ y $a_1 = a_2$
- For each player $i \in N$ a preference relation $\succeq_i$ defined over Z.

One way of visualizing this type of games is using a *tree*, in which each internal nodes are identified with non-terminal histories and each one of these corresponds to a decision made by a player. Given a history h in the game, which corresponds to a path in the tree, player $\rho(h)$ will choose on action from the set $\chi(h)$

In this model, each player knows how did he get up to certain point (knows what history occured). To allow more generality, we consider cases where each player doesn't know how he got up to certain point; i.e., he does not know which history happen and he can't distinguish in which node is he.

Since this model is about sequential decisions, it is useful for a player to define a plan of action for different scenarios, this is called a *strategy*; i.e., a plan that specifies the action chosen by each player for every history after which it is his turn to act.

3 Problem Statement

3.1 Problem Statement

Let $\mathcal{G}$ a causal graphical model and let $(\mathcal{A}, \mathcal{E}, \mathcal{C})$ a decision problem under uncertainty whose actions $a_i = \{c_j | E_j : j \in J\}$ are causally related to consequences $c \in \mathcal{C}$ through the uncertain events $E \in \mathcal{E}$ which correspond to variables in $\mathcal{G}$. Consider a rational decision maker who doesn't know the parameters of the causal model which control the probabilities of observing a consequence given an action $a \in \mathcal{A}$, we ask how the decision maker could learn about the causal structure that controls his environment in order to make good choices with respect to the decision makers' preferences.

3.2 Motivation

Decision making under uncertain conditions is a fundamental part of intelligent reasoning (Lake et al. (2017)). Intelligent agents often face situations where an action must be chosen in the presence of uncertain conditions; this means that an outcome will be observed according to some probability distribution given the action chosen by the agent.

In many real-world applications, the agent doesn't know all of the parameters required to calculate the maximum utility, but if the agent knew that his actions and the possible consequences were *causally related*, then he could attempt to discover this relations and use them in order to predict consequences of actions better than if he only observes multiple action-outcome pairs as done in Reinforcement Learning (Sutton and Barto (1998)).

It is known that human beings conceive their actions on the world as *intervening* in the world (Hagmayer and Sloman (2009)). Following this idea, Lattimore et al. (2016) consider decision problems where the action to be chosen is an intervention over a known causal graphical model. The agent must choose the intervention that maximizes the value of a *target variable* after a series of learning rounds. They model their problem as if choosing an intervention was choosing an arm of a *slot machine*, in which a gambler chooses an arm and gets some reward. From the rewards they estimate probabilities and output an optimal action in the sense of obtaining minimal regret. Their work considers that the causal model is fully known. They mention that the case where the causal model is unknown is left as an open question, and it is precisely what we are proposing to answer.

3.3 Justification

Many real-world applications of decision making are solved by *associative* methods which capture only statistical patterns that are found in data. For example, current methods in Reinforcement Learning, and specifically in Deep Reinforcement Learning, although they have good performance in the task that were supposed to solve, they can not explain *why* a specific trajectory was chosen by the algorithm. This is highly relevant in real-world applications such as self-driving cars, where it is very important to understand why an accident happened.

For example, as told by Bornstein (2016), at the University of Pittsburgh Medical Center a team of researchers tried to use Machine Learning to predict whether pneumonia patients

might develop severe complications. For this purpose they trained Neural Networks and Decision Trees using the hospital's own data. Neural Networks outperformed Decision Trees, but only by studying the decisions made by the latter did the doctors find out that the algorithms instructed doctors to send home pneumonia patients who already had asthma, despite the fact that asthma sufferers are known to be extremely vulnerable to complications. The problem relied in the training data, because the hospital policy was to automatically send asthma sufferers with pneumonia to intensive care, and this policy worked so well that asthma sufferers almost never developed severe complications. It was only through the interpretability of the Decision Trees that the doctors didn't send asthma patients with pneumonia home to a certain death.

Methods based in Deep Neural Networks aren't supposed to explain why a certain output was produced since those methods are based in *parallel distributed representations* and the same goes for any other learning algorithm that uses Deep Neural Networks, like Deep Reinforcement learning; when AlphaGo (Silver et al. (2017)) defeated the world-champion, humanity didn't learn anything new about the game of Go, because the algorithm was not designed to explain its moves, it was just curve fitting, although a very sofisticated method for achieving it. Using only associations between variables it is not possible to infer which one is the cause and which one the effect, something extra is required. The fundamental difference between causal and associative models is that causal models consider *doing* over *observing* (Pearl and Mackenzie (2018)). Once a causal relation has been established between two events, performing, or intervening, over a cause allows to predict the outcome in a more robust way than if using only correlations between observations.

On the other hand, learning a causal model of an environment and using it to act upon the environment allows to *explain* aspects of the model that a purely associative model would not be abel to explain. It allows to ask *why*.

The impact of Causal Inference goes beyond Machine Learning and Computer Science. For example, economists are interested in understanding what did certan public policy caused (Athey (2017)) or how human decision makers that show *inconsistent* preferences over time could be oriented if the causal consequences of this inconsistent behavior is introduced into a decision-making model (Albers and Kraft (2016)). In complex adaptive systems, causal relations can be used to clarify the complex interactions between agents in the system (Abbott and Hadžikadić (2017)) as well as for prediction and planning (Hunt et al. (2016), Brock (2018)).

3.4 Research Question

How a rational decision maker who faces an uncertain environment which is governed by a causal mechanism can learn and make use of this causal structure in order to make good choices? How can a causal structure help a decision maker in order to guide his learning process? What does the rationality assumption implies about how to choose when considering causal information? How to trade off exploration and exploitation when trying to learn about the causal structure of an environment while also trying to make good choices?

3.5 Hypothesis

Let $\mathcal{G}$ a causal graphical model and let $(\mathcal{A}, \mathcal{E}, \mathcal{C})$ a decision problem whose actions $a_i = \{c_j | E_j : j \in J\}$ are causally related to consequences $c \in \mathcal{C}$ through the uncertain events $E \in \mathcal{E}$ which correspond to variables in $\mathcal{G}$. Then, if a decision maker doesn't know the probabilities of the uncertain events in E, by repeatedly making decisions he can learn a causal model of the environment and use it to find the optimal action (in the sense of expected utility theory) in less, or equal, rounds than if he doesn't consider causal information.

3.6 Objectives

3.6.1 General Objectives

The proposed research has as a general objective to understand what are the implications of causality for rational decision maker who faces an uncertain environment and how causal relations can be discovered and used in order to make good choices that maximize the expected utility for the decision maker.

3.6.2 Specific Objectives

1. In order to achieve the general objective, we need to separate it into smaller problems which are required to be solved first. In first place, it is required to find a general way of defining a game that captures the interaction between a decision maker and a causal-governed environment.

2. In second place, to find a family of distributions that represent any past knowledge and current beliefs about the causal structure of the environment and a way to use these beliefs in order to make choices according to the decision maker's preferences.

3. Beliefs must be updated after external information is provided by the environment in response to any given action that was taken by the decision maker. A general updating criteria must be provided which uses the causal nature of the environment.

3.6.3 Limitations

The proposed research has some limitations. We are considering a *rational* decision maker whose preferences can be represented as if maximizing an expected utility with respect to a subjective probability distribution and an utility that assigns numbers to possible outcomes. Maximization of expected utility is not the only decision making criteria as other theories exist, such as Prospect Theory (Kahneman and Tversky (1979)) where a decision maker considers a reference point and from there he considers losses more painfull than gains. Other theory is Gilboa's Case-Based Decision Theory, where a decision maker considers how similar is the current decision problem with problems faced in the past.

We consider a particular definition of Causality, which is Spirtes' definition (Spirtes et al. (2000)) who defines causality as a stochastic relation between events. We take this definition

of causality as well as its theoretical requirements as axioms and we do not question if is the definition most suitable for a certain situation.

We are not considering that the player called Nature has any intentions nor objectives since we are given her constant payments in every outcome. A scenario where Nature could have objectives to pursue if it is modelling a player such as a Government, who can only act according to current laws but has general social objectives, such as the maximization of people with jobs in a negotiation with Unions.

3.6.4 Extensions

Given the Limitations addressed in the previous section, some possible extensions could be:

- Consider longer games.
- Consider more players than an agent and Nature.
- Consider objectives/intentions for Nature.
- Studying game-theoretical properties of the model such as equilibria in terms of the causal structure.

4 Related Work

Reinforcement Learning (RL) has been the standard setting for decision-making and learning by interaction. In this setting it is assumed that a Markov Decision Process (MDP) is able to capture the effects of the agent's actions in the environment. An MDP is composed by a set of states, a transition function and a reward function, where an agent's action moves the state of the environment according to the transition function and obtaining a reward in such process.

Reinforcement Learning can be roughly divided into two categories: model-based and model-free methods. In the former case the agent learns an *optimal policy* from a model of the environment, such as the Dynamic Programming method explained in Sutton and Barto (1998). On the other hand, if the agent does not know the transition function, then he must learn directly values of *state-action* pairs. A very well known model-free method is the classic Q-learning algorithm developed by Watkins and Dayan (1992). In either case, model-free or model-based methods an agent must use information from transitions or from observations. For this reason, Reinforcement Learning uses only associative patterns for finding an optimal policy and this methods can not learn beyond what is expressable by an MDP, which is constrained into associative rules and can not express any other kind of structure, such as causal relations.

Since an agent that learns how to act according to any Reinforcement Learning method acquires an optimal policy, optimal in the sense of satisfying the Bellman equations, it can be shown that such policies actually achieve the maximum expected utility as required by the rationality condition (Webb (2007)). In this way, when only associative information is available, RL is a coherent learning framework. On the other hand, it remains to be answered how to incorporate a causal model into an on-line decision making algorithm.

The formation and use of causal knowledge in human beings can provide inspiration for our problem, and it has been extensively studied by York Hagmayer and his team They have found that human beings, while facing a sequential decision problem, use and modify causal information (Hagmayer and Meder (2008), Hagmayer and Sloman (2009), Hagmayer et al. (2010), Hagmayer and Meder (2013)). In their studies, all of the subjects were faced with a one-stage decision problem in which an initial causal model was provided. In several rounds, the subject made a decision and observed a feedback and was later asked to describe the causal model they thought controlled the situation. The subjects knew that they were supposed to learn a causal model, but the case where the final objective is other than directly learning a causal model has not been studied. Bramley et al. (2015) show that people acquire causal structures by selecting the *most informative* options, in the sense of decreasing uncertainty about the true model instead of maximizing expected utility.

It is also known that human beings focus on *local* aspects while learning causal relations that are later unified into a single structure (Fernbach and Sloman (2009), Waldmann et al. (2008), Wellen and Danks (2012), Danks (2014)). Following this idea, Wellen and Danks (2012) propose a model to explain how observations and interventions are used by human beings to learn causal structure when little prior information is available.

An attempt to formalize Decision Theory in the presence of Causal Information was at-

tempted by Joyce, who stated that a decision maker must choose whatever action is more likely to (causally) produce desired outcomes (Joyce (1999), Peterson (2017)). His formulation falls short since he does not consider any kind of formal causal model beyond what is commonly understood by causality, although he captured the intuition that causal relations may be used to control the environment and to predict what is caused by the actions of a decision maker. In his formulation of a Causal Decision Problem, he uses the *probability of causing certain effect with an action*, which we refine here with the machinery of Pearl's Do-Calculus and Spirtes' definition of Causality.

In the class of decision problems considered for this proposal it is assumed that the a rational agent must choose some available action and then receive a reward from it. In the case where the agent does not has enough information available, we consider a certain number of learning rounds and expect the agent to output a good action. Problems of this nature have been modelled as *bandit* problems (Sutton and Barto (1998)). A bandit is an analogy of a *slot machine* as the ones found in Las Vegas, where a lever is pulled an some monetary reward (probably cero) is observed. A bandit problem consists by a set of available actions (the arms of the slot machine) and a reward function. It has been shown that the probability of the action selected in the n-th round not being optimal has an upper bound (Audibert and Bubeck (2010)) if the optimality criteria is choosen to be the *minimal regret* which is basically the difference of the theoretical mean reward of the optimal action and the chosen action. On the other hand, if the largest mean reward is desired with a fixed confidence of δ then an algorithm for finding the optimal action exist in such a way that the number of rounds to find it is within a constant factor of a lower bound (Jamieson et al. (2014)). If one does not consider a fixed confidence but instead a fixed number of learning rounds (known as fixed budget), then we face a *fixed budget* problem where a lower bound for the probability of finding the optimal arm is given by Carpentier and Locatelli (2016).

Notice that none of the previous works consider any kind of causal structure. For this kind of problems, causal relations between actions and consequences could be considered, in this way an action could be conceived as an *intervention* over some environment, which is in fact how humans consider actions in the world (Hagmayer and Sloman (2009)) and that human beings use and modify causal knowledge during a sequential decision making process (Hagmayer and Meder (2013)). Consider a set of variables that are causally related between them and to a *reward* variable whose value has to be maximized by intervening one of the variables. It is shown by Lattimore et al. (2016) that adding causal information in a fixed budget decision problem allows the decision maker to learn *faster* had he not considered causal information. Their work requires that the causal model is fully known to the decision maker, this requirement is relaxed later by Sen et al. (2017) who requires only that some part of the causal model is known and allow interventions over the unknown part.

In Lattimore et al. (2016) a causal graphical model $\mathcal{G}$ is assumed to be known and a number of learning rounds T is fixed. In round $t \in [1, ..., T]$ the decision maker chooses $a_t = do(X_t = x_t)$ and observes a reward Y_t. After the T learning rounds, the decision maker is expected to choose an optimal action a^* that minimizes the *expected regret*, which is defined as $R_T = \mu^* - \mathbb{E}[\mu_{a^*}]$ where $\mu^* = \max \mathbb{E}[a]$. They show that the achieved regret is of a smaller order than the regret obtained by non-causal algorithms.

Discovering the causal model itself while using current knowledge to make choices is left as

future work in both Lattimore et al. (2016) and Sen et al. (2017), but algorithms to discover causal relations in data can be found in Eberhardt (2008), Hauser and Bühlmann (2012), Hyttinen et al. (2013), Loh and Bühlmann (2014), Shanmugam et al. (2015) Mooij et al. (2016). Also, an *active learning* approach can be used to learn causal models from data, where one starts with some initial graph and then select the instances in the data that allow to add and orient edges in order to end with a fully oriented graph. Active learning algorithms for causal discovery can be found in Tong and Koller (2001),Murphy (2001), Meganck et al. (2006), He and Geng (2008), Hauser and Bühlmann (2012), Ness et al. (2017), Rubenstein et al. (2017).

Notice that these papers consider learning while interacting with a causal environment. We propose to model this interaction as a game, which has been previously considered in Werling et al. (2015), who consideres an *on-line* classification problem and modeled it as a game and where an oracle is used to classify initial observations.

5 Methodology

In this section we describe the steps that are to be followed in order to answer the Hypothesis.

5.1 A guiding principle for a Causal Decision Problem

We propose the following principle as a normative guide for how causal information must be used in order to find a solution to a decision problem where the preferences of the decision maker are assumed to be rational. We consider only the case where the decision maker's utility is given by the outcome of a $\{0, 1\}$-binary variable.

Proposition 5.1. Let $\mathcal{G}$ a causal graphical model and let $(\mathcal{A}, \mathcal{E}, \mathcal{C})$ a decision problem under uncertainty whose actions $a_i = \{c_j | E_j : j \in J\}$ are causally related to consequences $c \in \mathcal{C}$ through the uncertain events $E \in \mathcal{E}$ which correspond to variables in $\mathcal{G}$. Let Y a variable in $\mathcal{G}$ the variable whose realizations correspond to the consequences of the decision problem once the agent has chosen an action $a \in \mathcal{A}$ and assume that Y only takes values in the set $\{0, 1\}$ where 1 is a more desired outcome for the agent. Then, the *solution* for this decision problem is given by the action $a^* \in \mathcal{A}$ that satisfies:

$$P(Y = 1|do(a^*) \geq P(Y = 1|do(a)) \text{ for all } a \in \mathcal{A}.$$

5.2 Modelling a Decision Problem under Uncertainty as a Game

Consider a decision problem under uncertainty $(\mathcal{A}, \mathcal{E}, \mathcal{C}, \succeq)$ where the available actions $\mathcal{A}$ to a decision maker are causally related to the outcomes $\mathcal{C}$ and this causal relations are captured by a Causal Graphical Model $\mathcal{G}$. The Causal model is assumed to remain fixed and unknown to the decision maker, who is assumed to be rational and aware of the causal nature of the decision problem he faces.

As stated in the Hypothesis, we propose that the agent learns from the environment by interacting with it. By interaction we mean a number of rounds where the decision maker will act and then observe the outcome of his action.

Since the observed outcomes, given an action, are guided by the Causal Model, we can think of the environment as a decision maker whose available actions depend on what has been chosen previously by the original decision maker. We have mentioned that Game Theory allows to model situations where two or more decision makers interact. In this way, the first step in order to solve the stated problem is to model the interaction of the (original) decision maker and his environment as a **game** between two players: the decision maker, and a player which will be called Nature, whose actions are to be guided by the causal model. Since in decision problems under uncertainty it is assumed that the *state* of the environment is unknown, we will assume that player Nature has the first move and he assigns some state to any of the variable in the causal model $\mathcal{G}$.

After the decision maker makes his *play*, Nature will *respond* to the action selected according to the causal relations expressed by $\mathcal{G}$. This action-response dynamic forces to consider

some notion of order between the players' moves, and because of this reason we must use the *extensive-form* games described in Section 2.6.1.

We are considering each game as a learning round, where the only information that is carried from a game to another is what is learned by the decision maker via a process of belief updating. The causal model is assumed to remain fixed and each initialization at the beginning of each new game is independent from any past outcome.

The extensive-form game considers two possibilities: perfect or imperfect information, where in the latter a player may not know what actions have been played in the past by other players, which is what we need to use because the decision maker does not know what move has Nature selected.

5.2.1 Elements of the Game

The game which will be used to model the interaction between a decision maker and his environment is composed by:

- **Players:** The decision maker and a player called Nature, who will move first in order to assign a state.
- **Actions:** The actions available to the decision maker are the same ones as in the original decision problem. Nature must choose according to the causal model.
- **Preferences:** The preference relation to the decision maker is his same rational preference relation as in the decision problem. Nature is indifferente between outcomes.

5.3 Belief Formation

Once a Game that models agent-environment interaction is defined, we must turn to the question about *learning* about the environment, which includes acquiring causal information and also using current causal knowledge in order to obtain new information.

For the agent to learn and reason about the plays of Nature we assume that he is able to observe the final outcome of the game, which includes knowing about the inicial uncertain state that Nature assigned. The agent, at the beginning of the game is uncertain about the state of the environment which he is facing, but he still must take an action, so in order to allow the decision maker to use his knowledge to act, he will have *beliefs* about any relevant aspects of the environment which he will use as if they were the true model in any given play. These beliefs will be *updated* according to what is observed at the end of each game. The second step is to find a way to specify beliefs about causal structure.

Beliefs represent ignorance about relevant characteristics of certain environment (Bernardo (2000), Peterson (2017)), and thus are encoded as probability distributions, which usually come from a parametric family.

In our proposed research, beliefs will encode current knowledge about the *causal structure* of the environment, and this beliefs will be updated when information about actions-outcomes becomes available, which will happen at the end of each game. The representation of the

beliefs must allow them to be used in order to make a choice using the causal information they represent in each round.

Beliefs must also encode any previous knowledge possesed by the decision maker, and this initial assignment must be coherent with the actions prescribed with the rationality axioms for preferences. In Billot et al. (2005) is shown how to build a set of beliefs given a set of past observations in an axiomatic way.

5.4 Belief updating

The whole idea is that a decision maker is able to learn from his environment by observing, and reasoning about, the effects of his actions. The observed information will modify his current knowledge, which is encoded as probabilistic beliefs about the environment in a way that is similar to how human beings modify causal knowledge while intervening in the world (Hagmayer and Meder (2013)).

If a decision maker encodes his current knowledge, and ignorance, about some relevant characteristic of his environment, then it is known that belief updating using Bayes' Theorem is the only way of updating that is coherent with the rational preference axioms (Bernardo (2000)).

Bayesian updating is also of great importance in the theory of Learning in Games (Fudenberg and Levine (1998)) since it is known (Shoham and Leyton-Brown (2008)) that if all the players in a game perfom Bayesian updating, then the plays generated by these measures converge to the plays that would be generated if all players shared their knowledge.

Even though we know that Bayesian updating is the correct *framework* for belief updating, it remains unanswered how to concretely use causal information in order to update the parameters that control the distributions used to express beliefs in a tractable way. Therefore, the third step is to find an updating criteria with theoretical properties that guarantee convergence to the true causal model that controls the environment. The updating criteria must explicitly use causal information beyond any associative rules.

5.5 Solving first simpler cases

Once that the model has been established, we proceed to further examine the problem of acquiring and using causal information in a decision problem and we notice three important cases, in order of ascending difficulty. In all three cases we consider a decision maker who faces an uncertain environment and who must take an action in order to satisfy his preferences. The uncertain environment in which the decision maker (or agent) exist is governed by a causal graphical model $\mathcal{G}$ which relates actions taken by the decision maker with outcomes. We are assuming that one of the variables of the casusal model is of interest for the decision maker in the sense that he seeks to maximize its value. This variable will be known as *target variable* or reward variable.

The three cases that we notice are:

- The decision maker knows the causal model.
- The decision maker knows only the graphical *structure* of the causal model.

- The decision maker knows nothing about the causal model.

In the following we show how the first two cases, which are particular cases of our general problem, can be solved and we show that the proposed solution for the first two cases sheds light on how the general case could be solved.

5.5.1 Fully knowing the causal graphical model

If the decision maker knows the causal graphical model, then following the guiding principle, he could easily calculate the effect on the target variable that any of his actions has and then choose the action that has the highest probability of causing a *desired* value of the target variable. The decision thus made is the one that maximizes the expected utility, by construction. Also, choosing this action is the *best response* action, and therefore a Nash equilibrium for the game. Choosing the action that is more likely to produce the most desired outcome according to current causal knowledge is required by Joyce's Causal Decision Theory (Joyce (1999)).

5.5.2 Knowing only the graph structure

If a decision maker knows the structure of the graph, but he does not know the parameters, then he encodes this ignorance, along with any useful previous knowledge, as a probability distribution over a suitable space. In the preliminary results that will be shown later a particular case is considered, where the beliefs have the form of a Dirichlet distribution for each of the Conditional Probability Tables that characterize a Bayesian Network.

In general, what must be specified is a way of encoding current causal knowledge into a probability distribution in such a way that a *local* (i.e. in each round) causal model can be obtained from current beliefs in order to make a decision using current causal knowledge and those beliefs will be later updated using what is observed after choosing an action. The action will be chosen, given a set of beliefs, as in the previous case and by following the guiding principle. This means that using current beliefs, the agent forms a causal model which is used as if it were the true causal model.

5.5.3 General case: Model unknown

The previous two cases are per se steps of the methodology, rather simplifications in order to understand what is happening. The real deal is the case where the causal model is unknown.

When the causal model is unknown, two cases of interest appear, depending if the decision maker knows the (maximum) number of variables in the model or not.

If the maximum number of variables is assumed to be known by the decision maker, then the situation simplifies since a decision maker could generate distributions that represent causal relations among variables. For example, a Dirichlet Process (Ferguson (1973), Ghosal and van der Vaart (2017)) could be used to generate Dirichlet distributions that are used generate Conditional Probability Tables in order to specify a causal model as in the previous case.

If the number of variables is unknown, then the beliefs that the decision maker holds must allow unbounded cardinality on the number of variables. In that case, a distribution over a *graph space* could be used in such a way that when sampled a graph structure is obtained and used as in the previous case.

5.6 Balancing exploitationd and exploration

When the model is to be discovered, such as in the general case mentioned before, an agent must choose also between keep making an intervention in a variable that is known to cause a desired output or to explore for new variables that could bring even better rewards, but at the cost of not having best performance while looking out for them; this is known as the exploration vs exploitation dilemma (Sutton and Barto (1998)). This question is to be attacked using the concept of *mixed strategy* in a game. A mixed strategy is a probability distribution over available actions for an agent, in this way, similar to the ε-greedy methods in Reinforcement Learning, we allow for some actions to be randomly choosen in order to explore further possibilities.

5.7 Validation of the proposed methods

In order to show that an agent has succeeded in acquiring a causal model from his environment via interacting with it and using it in order to make good choices, it is necessary to compare the average utility obtained in a series of rounds with the theoretical utility that would be obtained if the causal model was fully known, and to compare the performance obtained by a learning agent that is not considering causal information. Also, the parameters of the learned causal model must be close to the true ones.

5.8 Experimental methodology

The proposed methodology will be tested in the following way: A fully specified causal model, which satisfies the conditiones required above, will be given and used as a ground truth. The causal model will not be available to the decision maker, since one of the basic assuptions is that the agent does not know the causal model that controls his environment. The true causal model will be used to simulate realizations of the other variables in the distribution given any action taken by the decision maker.

One, or more, of the variables is to be selected as *available* to interventions by the decision maker, the set of possible values for the selected variable determines the set of available actions of the decision maker in the formulation as a game. Also, another variable is selected as a *target* variable, preferably an endogenous variable in the model in order to be affected by any action taken by the agent. The set of possible values of this variable must be in correspondance with the set of outcomes of the decision problem that the agent faces so the value that the target variable outputs has an effect over the preferences of the agent.

A distribution with initial parameters θ_0 will be specified.

Given a number T of rounds, and inside round $t \in \{1, 2, ..., T\}$ the agent will make an action $a \in \mathcal{A}$ and then observe the state of the other variables of the model. Using this

information, current beliefs will be updated according to the criteria to be established.

The chosen actions will be compared to classic on-line decision making

5.8.1 Internal Validation: Reproducibility

For the results obtained by following the methodology described in previous sections be reproducible, the causal model used to simulate an environment must be fully specified by describing the probabilistic parameters that control it. The initial distributions given to the decision maker to encode his beliefs must be shown as well with the updating criteria used when information from the environment becomes available.

5.8.2 External Validation: Different problems

For the proposed method to be tested in different decision problems, it is required that the correspondent causal model satisfies all of the conditions previously described, and any derived algorithm will be designed in such a way that the characteristics of the causal model are an external input. In this way, if an external user is able to give a causal model, the method is expected to work independent of the particularities of the causal model which is being tested.

5.8.3 Concurrent Validation

Concurrent validation will be tested by comparing the performance obtained by the proposed method against the theoretical optima.

5.9 Working plan

Concrete steps towards solving the problem:

- Literature review, understanding the problem, considering several models and possible solutions.
- Studying implications of the rationality assumption for decision making under uncertainty in causal environments.
- Formalization of a Causal Decision Problem: a decision problem under uncertainty where causal relations are encoded into a Causal Graphical Model. Propose a solution using the rationality assumption.
- Propose a representation of beliefs that allow causal structure to be considered. Three subcases considered
- Propose and evaluate an updating criteria for causal beliefs that can be updated according to new information.
- Study the problem of on-line causal discovery.
- Considering interventions in more than one variable.

5.9.1 Publications plan

- **Year 1:** Gonzalez-Soto, M., Sucar, L.E., Escalante, H.J., **Playing against Nature: causal discovery for decision making under uncertainty**. Accepted for poster presentation at the **CausalML 2018 Workshop at ICML 2018**. July 15 2018, Stockholm, Sweden. This can be confirmed at `https://sites.google.com/site/faim18wscausalml/program-schedule`. The paper can be found at `https://arxiv.org/abs/1807.01268`.

- **Year 2:** Conference Paper

  - The implications of Causal Relations for Decision Making: a general principle for Decision under Uncertainty.

- **Year 2:** Conference Paper

  - When a causal model is unknown, how to encode uncertainties about causal structure in a probabilistic distribution? Bayesian Nonparametric prior that explicitly represents causal structure.

- **Year 3:** Journal Paper

  - Aim: Theoretical paper of results for on-line causal learning.

6 Preliminary Results

To show the factibility of this proposal, we considered a test scenario and two cases in ascending difficulty.

6.1 Test scenario

Consider a sick patient who arrives at a hospital and he either has disease A or disease B. The doctor can either give him some pill or send him into surgery. Both treatments entail risks and whether the treatment cures the patient or not depends on which disease it had originally. The doctor could be facing a mutation from a known disease, so she has some knowledge about what could happen if a treatment is given to the patient. Using her previous knowledge as a true model, she can choose a treatment and observe the outcome from which she will learn about this disease, so she could make a better decision the next time a similar patient arrives.

The causal model that governs this situation is shown in Figure 1. The parameters for this model were fixed intuitively in such a way that each treatment is effective for only one disease, but the most effective treatment is riskier.

The variables in the model are:

- **Disease:** Either A or B.
- **Treatment:** Either pill or surgery.
- **Reaction:** Either dying or surviving.
- **Lives:** Either living or dying.

The variables are causally related as shown in Figure 1.

The variable *Lives* is the *target variable* and, in this example, the only variable that can be intervened upon is the variable *Treatment*. The decision maker prefers an outcome in which the patient lives.

In this scenario, Nature's move will consist in randomly assigning a disease to the patient. Then, the medic will asign a treatment using his current beliefs about the disease and the possible outcomes. The decision nodes for this play of the medic form an *information set* because the medic doesn't know how she arrived there since she doesn't know what disease did Nature assign. Finally, Nature will sample the consequence of the treatment from the causal model and the medic will observe the outcome.

For this test scenario whose causal graphical model is shown in Figure 1, we see by applying the Pearl's do-calculus that the interventional distribution $P_{do(Tr)}(Y)$ is given by

$$P(Y|do(Tr)) = P(Y|D, Tr, R)P(R|Tr)P(D).$$

In fact, from the structure of the model, which is shown in Figure 1, we see that the involved

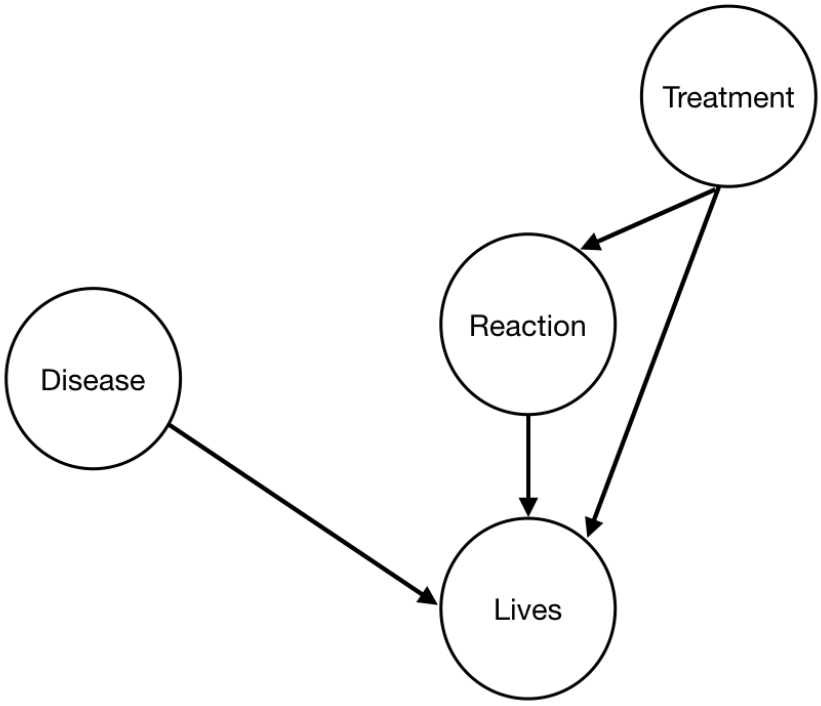

Figure 1: Causal graphical model for the test scenario: the target variable *Lives* is causally influenced by the disease the patient has, the treatment assigned and the survival to the secondary effects of treatment.

probabilities in any calculations are:

$$P(\text{disease}), P(\text{treatment}), P(\text{reaction}|\text{treatment}),$$

and

$$P(\text{lives}|\text{disease, treatment, reaction}).$$

We can also see that the joint distribution for all of the variables can be expressed as

$$P(Y|D, Tr, R)P(R|Tr)P(D)P(Tr).$$

This expression will be useful when specifying beliefs about the model as probability distributions.

6.2 Case 1: The causal model is completely known

If the causal model is completely known to the decision maker, then in one step she can obtain the probability for her desired value of the target variable, which in this case is the value

corresponding to the outcome in which the patient lives at the end. Using this probability, she can choose which treatment to assign. Since this action maximizes the probability of the occurence desired value, it maximizes the expected utility, and it is also a *best response* to the player Nature.

6.3 Case 2: Only the structure is known

Since the decision maker knows the graph structure, he can explicitly find a non interventional expression for the interventional distribution and update his beliefs about these unknown quantities. If the decision maker were not allowed to know, at the end of each round, the play of the Nature then this will have to be estimated as a hidden variable using, for example, the EM algorithm Dempster et al. (1977), but meanwhile we are assuming that this information is available at the end of each round.

Given the structure of the model; i.e., the variables in it and the directed edges, the joint distribution of those variables can be expressed as a product of the form $P(X_j|Pa(X_j))$ where $Pa(X_j)$ are the parents of X_j in the underlying DAG in $\mathcal{G}$. Since these distributions fully characterize the model, the decision maker will have beliefs over each one of these parameters. Notice that each of these parameters is itself a distribution of length equal to the number of possible values of the variable which is being conditioned, call the maximum number of possible values k .

A distribution suitable to modelling discrete probability vectors is the k-dimensional Dirichlet distribution, whose support is the set of probability vectors of length k Hjort et al. (2010). The k dimensional Dirichlet distribution has a density f with respect to the Lebesgue measure given by

$$f(x_1, ..., x_k|\alpha_i, ..., \alpha_k) = \frac{1}{B(\alpha)} \prod_{i=1}^{k} x_i^{\alpha_i - 1},$$

where $(x_1, ..., x_k)$ are such that $\sum_{i=1}^{k} x_i = 1$ and $\alpha = (\alpha_1, ..., \alpha_k)$. The Dirichlet distribution is useful since it is conjugate for itself Bernardo (2000).

The decision maker will have beliefs about the CPT's in the form of parameters of several Dirichlet distributions. Using the agent's current beliefs, a causal graphical model can be specified. Using this fully specified (structure + parameters) as a true model, the decision maker will make his choice as in Case 1. When the decision maker observes the value of the target variable, he will update the parameters that specify his beliefs.

Previously we argued that the agent's beliefs were going to be *distributions* over a suitable space, but what is going to be updated are the parameters of such distributions. Namely, the α corresponding to the Dirichlet random variable assigned to each CPT.

For the belief updating, given a new data point, two cases must be considered:

- The variable to update has no parents.
- The variable to update has parents.

In the first case, if a prior Dirichlet(α) is used, then the posterior is given by

$$\text{Dirichlet}(\alpha + c)$$

where c is a vector of the number of occurrences of that observed data point.

For the second case, we must consider both the occurrences of that data point as well as the parents for each of the variables. Following Barber (2012) we denote as $\theta_i(X, j)$ the number of times the event $\{X = i|Pa(X) = j\}$ is observed. In this case, if the prior of X_i conditioned on its parents having the value j is given by a a Dirichlet(α), then the posterior for the variable X_i given an observed data point is given by

$$\prod_j \text{Dirichlet}(\alpha + \theta_i(\cdot, j)).$$

6.3.1 Implementation

We begin with a random assignation of the α parameter for each of the distributions considered. We use Dirichlet distribution for each of the conditional probability tables that appear in the factorization of the joint probability for the graph of $\mathcal{G}$. Since each of the variables in the model is binary, then the product of these Dirichlets is Dirichlet.

With this parameters, the decision maker forms a causal model and chooses the action that maximizes the probability of the desired value for target variable as in Case 1. With this action chosen, we simulate an outcome from the causal graphical model using the chosen action as an intervention. This evidence is used to update the parameters, which then will be used to generate a new causal model, and so on.

We show the results of two experiments. We compare the performance obtained by the causal agent, a *random agent* who selects his actions at random, and an agent performing Q-learning (Watkins and Dayan (1992)). Q-learning was chosen since it learns an *optimal policy* in the sense of the Bellman Equation (Sutton and Barto (1998)) and it is shown in Webb (2007) that such optimal policies also maximize expected utility.

The average perform over $20, 50, 100$ and 200 rounds are shown.

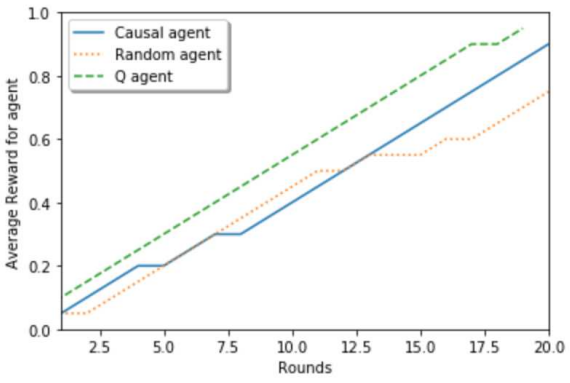

Figure 2: Average reward in 20 rounds

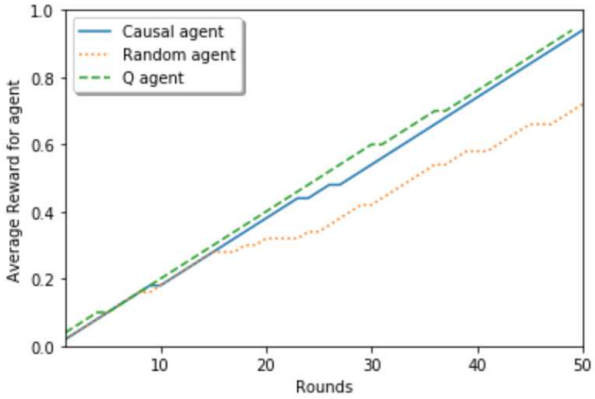

Figure 3: Average reward in 50 rounds

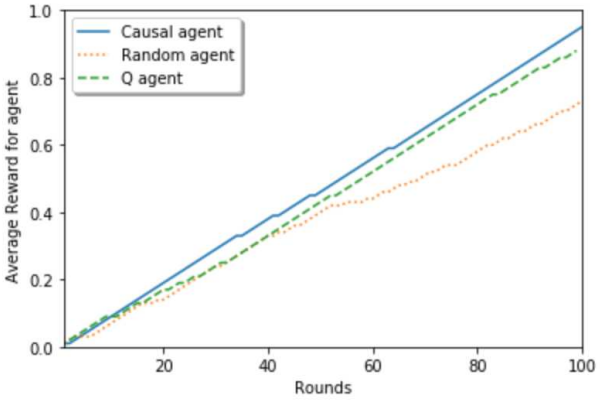

Figure 4: Average reward in 100 rounds

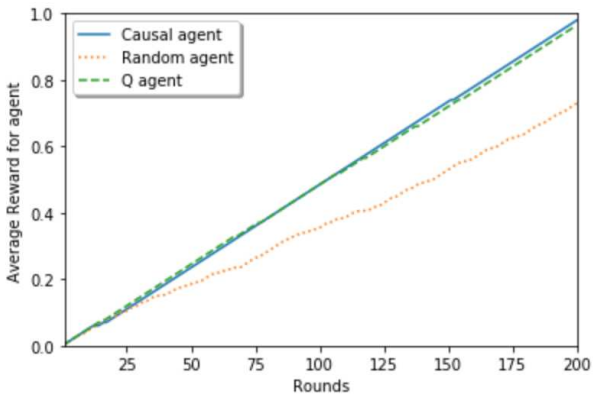

Figure 5: Average reward in 200 rounds

6.4 Case 3 and future work: The model is not known

The causal model were fully unknown, the decision maker will have to deal with the problem using only any previous knowledge and her own intuitions. Again, any previous knowledge and considerations will be expressed as *beliefs* about the uncertainties in the environment, which will take the form of a probability distributions over a suitable space.

As in the previous case, we consider a repeated game where the base game consists of Nature assigning a random state of the environment and responding to the agents choices with the effects that were caused by her decisions. In this game, as well as in the previous one, the decision maker will attempt to learn by updating, and using, beliefs in a suitable way.

The most notable difference with the previous case is that the *structure* of the model is also to be learned in such a way that both the structure and parameters converge to the true model in the limit. In the previous case the decision maker knew the form of the Conditional Probability Tables (CPT) involved in any calculation. In this case, she doesn't know the structure of the DAG so which CPT's are involved is unknown.

If the decision maker knew which variables appear in the true model that governs the environment, even though she didn't know how they are connected, she could use a *Dirichlet Process* to generate Dirichlet distributions and generate causal graphical models the same way as in Case 2 and updating the parameters of the process using the observed information. The Dirichlet Process[6], which was introduced by Ferguson (1973), is random measure defined over a space S such that for each partition $B_1, ..., B_k$ the vector $(G(B_1), ..., G(B_k))$ follows a Dirichlet distribution Hjort et al. (2010), Müller et al. (2016), Ghosal and van der Vaart (2017).

Belief updating using causal information when the decision maker doesn't know the structure of the model nor its parameters is yet to be studied and left as future work.

[6]with parameters M, G_0

References

Abbott, R. and Hadžikadić, M. (2017). Complex adaptive systems, systems thinking, and agent-based modeling. In *Advanced Technologies, Systems, and Applications*, pages 1–8. Springer.

Albers, S. and Kraft, D. (2016). Motivating time-inconsistent agents: A computational approach. In *International Conference on Web and Internet Economics*, pages 309–323. Springer.

Ash, R. B. and Doleans-Dade, C. (2000). *Probability and measure theory*. Academic Press.

Athey, S. (2017). The impact of machine learning on economics. In *Economics of Artificial Intelligence*. University of Chicago Press.

Audibert, J.-Y. and Bubeck, S. (2010). Best arm identification in multi-armed bandits. In *COLT-23th Conference on Learning Theory-2010*, pages 13–p.

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge University Press.

Bernardo, J. (2000). Bayesian theory. *Wiley Series in Probability and Statistics.*

Billot, A., Gilboa, I., Samet, D., and Schmeidler, D. (2005). Probabilities as similarity-weighted frequencies. *Econometrica*, 73(4):1125–1136.

Binmore, K. (2007). *Playing for real: a text on game theory*. Oxford university press.

Binmore, K. (2008). *Rational decisions*. Princeton University Press.

Bornstein, A. M. (2016). Is artificial intelligence permanently inscrutable? *Nautilus Magazine*, Issue 040, Chapter 1.

Bramley, N. R., Lagnado, D. A., and Speekenbrink, M. (2015). Conservative forgetful scholars: How people learn causal structure through sequences of interventions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 41(3):708.

Brock, W. A. (2018). Causality, chaos, explanation and prediction in economics and finance. In *Beyond Belief*, pages 230–279. CRC Press.

Carpentier, A. and Locatelli, A. (2016). Tight (lower) bounds for the fixed budget best arm identification bandit problem. In *Conference on Learning Theory*, pages 590–604.

Cartwright, N. (1983). How the laws of physics lie.

Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. MIT Press.

De Finetti, B. (1930). Funzione caratteristica di un fenomeno aleatorio. In *Atti della Academia Nazionale dei Lincei Rendiconti, Class di Scienze Fisiche, Mathematice e Naturali*, volume 4, pages 86–133.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

Eberhardt, F. (2008). Almost optimal intervention sets for causal discovery. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, pages 161–168. AUAI Press.

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1(2):209–230.

Fernbach, P. M. and Sloman, S. A. (2009). Causal learning with local computations. *Journal of experimental psychology: Learning, memory, and cognition*, 35(3):678.

Fudenberg, D. and Levine, D. K. (1998). *The theory of learning in games*. MIT press.

Garnelo, M., Arulkumaran, K., and Shanahan, M. (2016). Towards deep symbolic reinforcement learning. *arXiv preprint arXiv:1609.05518*.

Ghosal, S. and van der Vaart, A. (2017). *Fundamentals of nonparametric Bayesian inference*, volume 44. Cambridge University Press.

Gilboa, I. (2009). *Theory of Decision under Uncertainty*. Cambridge University Press.

Gilboa, I. and Schmeidler, D. (1995). Case-based decision theory. *The Quarterly Journal of Economics*, 110(3):605–639.

Glymour, C. (2003). Learning, prediction and causal bayes nets. *Trends in cognitive sciences*, 7(1):43–48.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, pages 424–438.

Hagmayer, Y. and Meder, B. (2008). Causal learning through repeated decision making. In *Proceedings of the Cognitive Science Society*, volume 30.

Hagmayer, Y. and Meder, B. (2013). Repeated causal decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(1):33.

Hagmayer, Y., Meder, B., Osman, M., Mangold, S., and Lagnado, D. (2010). Spontaneous causal learning while controlling a dynamic system. *The Open Psychology Journal*, 3:145–162.

Hagmayer, Y. and Sloman, S. A. (2009). Decision makers conceive of their choices as interventions. *Journal of Experimental Psychology: General*, 138(1):22.

Harnack, D., Laminski, E., Schünemann, M., and Pawelzik, K. R. (2017). Topological causality in dynamical systems. *Physical review letters*, 119(9):098301.

Hauser, A. and Bühlmann, P. (2012). Two optimal strategies for active learning of causal models from interventions. In *Proceedings of the 6th European Workshop on Probabilistic Graphical Models*, pages 123–130.

He, Y.-B. and Geng, Z. (2008). Active learning of causal networks with intervention experiments and optimal designs. *Journal of Machine Learning Research*, 9(Nov):2523–2547.

Hitchcock, C. (2018). Probabilistic causation. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2018 edition.

Hjort, N. L., Holmes, C., Müller, P., and Walker, S. G. (2010). *Bayesian nonparametrics*, volume 28. Cambridge University Press.

Holland, P. W., Glymour, C., and Granger, C. (1985). Statistics and causal inference. *ETS Research Report Series*, 1985(2).

Huang, Y. and Valtorta, M. (2006). Pearl's calculus of intervention is complete. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, pages 217–224. AUAI Press.

Hunt, E. R., Baddeley, R. J., Worley, A., Sendova-Franks, A. B., and Franks, N. R. (2016). Ants determine their next move at rest: motor planning and causality in complex systems. *Royal Society open science*, 3(1):150534.

Hyttinen, A., Eberhardt, F., and Hoyer, P. O. (2013). Experiment selection for causal discovery. *The Journal of Machine Learning Research*, 14(1):3041–3071.

Jamieson, K., Malloy, M., Nowak, R., and Bubeck, S. (2014). lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439.

Joyce, J. M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Lake, B. M., Ullman, T. D., Tenenbaum, J. B., and Gershman, S. J. (2017). Building machines that learn and think like people. *Behavioral and Brain Sciences*, 40.

Lamport, L. (1978). Time, clocks, and the ordering of events in a distributed system. *Communications of the ACM*, 21(7):558–565.

Lattimore, F., Lattimore, T., and Reid, M. D. (2016). Causal bandits: Learning good interventions via causal inference. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems 29*, pages 1181–1189. Curran Associates, Inc.

Loh, P.-L. and Bühlmann, P. (2014). High-dimensional learning of linear causal networks via inverse covariance estimation. *Journal of Machine Learning Research*, 15(1):3065–3105.

Meder, B., Gerstenberg, T., Hagmayer, Y., and Waldmann, M. R. (2010). Observing and intervening: Rational and heuristic models of causal decision making. *The Open Psychology Journal*, 3:119–135.

Meganck, S., Leray, P., and Manderick, B. (2006). Learning causal bayesian networks from observations and experiments: A decision theoretic approach. In *International Conference on Modeling Decisions for Artificial Intelligence*, pages 58–69. Springer.

Mooij, J. M., Peters, J., Janzing, D., Zscheischler, J., and Schölkopf, B. (2016). Distinguishing cause from effect using observational data: methods and benchmarks. *The Journal of Machine Learning Research*, 17(1):1103–1204.

Müller, P., Quintana, F. A., Jara, A., and Hanson, T. (2016). *Bayesian nonparametric data analysis*. Springer series in Statistics.

Murphy, K. P. (2001). Active learning of causal bayes net structure.

Ness, R. O., Sachs, K., Mallick, P., and Vitek, O. (2017). A bayesian active learning experimental design for inferring signaling networks. In Sahinalp, S. C., editor, *Research in Computational Molecular Biology*, pages 134–156, Cham. Springer International Publishing.

Nichols, W. and Danks, D. (2007). Decision making using learned causal structures. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 29.

Osborne, M. J. and Rubinstein, A. (1994). *A course in game theory*. MIT press.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Pearl, J. (2009). *Causality*. Cambridge university press.

Pearl, J. (2018a). Theoretical impediments to machine learning with seven sparks from the causal revolution. *arXiv preprint arXiv:1801.04016*.

Pearl, J. and Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Pearl, J., M. D. (2018b). *The Book of Why: The New Science of Cause and Effect*. Basic Books.

Peters, J., Janzing, D., and Schölkopf, B. (2017). *Elements of causal inference: foundations and learning algorithms*. MIT Press.

Peters, J., Mooij, J., Janzing, D., and Schölkopf, B. (2012). Identifiability of causal graphs using functional models. *arXiv preprint arXiv:1202.3757*.

Peterson, M. (2017). *An introduction to decision theory*. Cambridge University Press.

Rottman, B. M. and Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological bulletin*, 140(1):109.

Rubenstein, P. K., Tolstikhin, I., Hennig, P., and Schoelkopf, B. (2017). Probabilistic active learning of functions in structural causal models. *arXiv preprint arXiv:1706.10234*.

Savage, L. (1954). *The Foundations of Statistics*. New York: John Wiley & Sons.

Sen, R., Shanmugam, K., Dimakis, A. G., and Shakkottai, S. (2017). Identifying best interventions through online importance sampling. In *International Conference on Machine Learning*, pages 3057–3066.

Shanmugam, K., Kocaoglu, M., Dimakis, A. G., and Vishwanath, S. (2015). Learning causal graphs with small interventions. In *Advances in Neural Information Processing Systems*, pages 3195–3203.

Shoham, Y. and Leyton-Brown, K. (2008). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

Shpitser, I. and Pearl, J. (2006). Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, pages 1219–1226. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999.

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., et al. (2017). Mastering the game of go without human knowledge. *Nature*, 550(7676):354.

Sloman, S. A. and Hagmayer, Y. (2006). The causal psycho-logic of choice. *Trends in Cognitive Sciences*, 10(9):407–412.

Sorin, S. (2003). Stochastic games with incomplete information. In Neyman, A. and Sorin, S., editors, *Stochastic Games and Applications*, pages 375–395, Dordrecht. Springer Netherlands.

Spirtes, P., Glymour, C. N., and Scheines, R. (2000). *Causation, prediction, and search*. MIT press.

Spohn, W. (2012). *The laws of belief: Ranking theory and its philosophical applications*. Oxford University Press.

Sucar, L. E. (2015). Probabilistic graphical models. *Advances in Computer Vision and Pattern Recognition. London: Springer London. doi*, 10:978–1.

Suppes, P. (1970). *A probabilistic theory of causality*. North-Holland Publishing Company Amsterdam.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Tian, J. (2002). Studies in caual reasoning and learning. Phd thesis, Department of Computer Science, University of California Los Angeles.

Tong, S. and Koller, D. (2001). Active learning for structure in bayesian networks. In *International joint conference on artificial intelligence*, volume 17, pages 863–869. LAWRENCE ERLBAUM ASSOCIATES LTD.

Von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Wakker, P. P. (2010). *Prospect theory: For risk and ambiguity*. Cambridge university press.

Waldmann, M. R., Cheng, P. W., Hagmayer, Y., and Blaisdell, A. P. (2008). Causal learning in rats and humans: A minimal rational model. *The probabilistic mind. Prospects for Bayesian cognitive science*, pages 453–484.

Watkins, C. J. and Dayan, P. (1992). Q-learning. *Machine learning*, 8(3-4):279–292.

Webb, J. N. (2007). *Game theory: Decisions, Interaction and Evolution*. Springer Undergraduate Mathematics Series.

Wellen, S. and Danks, D. (2012). Learning causal structure through local prediction-error learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 34.

Werling, K., Chaganty, A. T., Liang, P. S., and Manning, C. D. (2015). On-the-job learning with bayesian decision theory. In *Advances in Neural Information Processing Systems*, pages 3465–3473.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford University Press.