Article

Causal Economic Machine Learning (CEML)

Andrew Horton *

Article

Causal Economic Machine Learning (CEML)

Andrew Horton

Independent Researcher. BA Honours, Western University. MBA, Ivey Business School, 305 Sunnyside Crescent, London, Ontario, Canada; ahorton.mba2000@ivey.ca; 647–966–9870

Abstract: This paper proposes causal economic machine learning (CEML) as a research agenda that utilizes causal machine learning (CML) built on causal economics (CE) decision theory. Causal economics is better suited for use in machine learning optimization than expected utility theory (EUT) and behavioral economics (BE) based on its central feature of causal coupling (CC), which models decisions as requiring upfront costs, some certain and some uncertain, in anticipation of future, uncertain causally-linked benefits. This multi-period, causal-linked process incorporating certainty and uncertainty replaces the single period lottery outcomes augmented with intertemporal discounting used in EUT and BE, providing a more realistic framework for AI machine learning modelling and real world application. It is mathematically demonstrated that EUT and BE are constrained versions of CE. Causal coupling can also be applied at the macroeconomic level to gauge the effectiveness of policies that deliver various levels of cost and benefit coupling for individuals. With the growing interest in natural experiments in statistics and CML across many fields, such as healthcare, economics and business, there is a large potential opportunity to run AI models on CE foundations and compare results to models based on traditional decision making models that focus only on rationality, bounded to various degrees.

1. Introduction

Artificial intelligence (AI) involves the creation of machines that are able to think and act like humans. An important goal is to ensure that these are safe for humans and are consistent with human values (Gabriel & Ghazavi, 2021). As a result, the AI Alignment problem (Chaturvedi, 2023) is a central area of research in the science of AI. Getting machines to think like human brains requires modelling machine learning in line with the way human brains make decisions in the real world (Sejnowski, 2018). The highly formalized human decision making models of economics based on utility functions lend well to this purpose. In the science of AI, the counterparts to economic utility functions are known by various terms, including value functions, objective functions, loss functions, reward functions and preference orderings (Naudé, 2023).

Explicitly grounding AI in economic theory has been shown to yield improved results (Jenkins et al., 2021). Many AI scientists strive to make their AI systems fully rational by utilizing rational choice theory from economics (Green, 2022), specifically in the form of expected utility theory (Naudé, 2023). But there is much evidence that human decision makers are not 100% rational and calculating in practice (Dillon, 1998). It has also been well documented that expected utility is not effective as an applied model of individual choice (Tversky, 1975). Comprehensive reviews of experimental research have demonstrated violations in each of the axioms of expected utility (Yaqub et al., 2009). Behavioral economics (Dhami, 2016) emerged as a powerful approach to decision making heavily grounded in psychology, incorporating the evidence that humans are restricted to bounded rationality in practice (Simon, 1990). The reality of decision making reflects many challenges—including computation, information and biases. Behavioral economics has been most robustly

structured within the framework of cumulative prospect theory (Tversky & Kahneman, 1992) and this has been a big leap forward in modeling. Today, much of AI modelling falls back on some form of bounded rationality (Şimşek, 2020).

These are significant enhancements in theoretical foundations, yet when it comes to real world application in decision making under uncertainty, economics and AI both face an environment with multiple individuals, human agent collectives (Naudé, 2023) and decisions that span multiple, sequential, causally-linked time periods where information is costly (Naudé, 2023). AI is a time-based process, with causality from percepts to outcomes (Russell & Norvig, 2021). Given this context, the theoretical frameworks of EUT and BE typically have to be put aside in applied economics in favour of dynamic stochastic programming (Bellman, 1957) based on a Markov decision process (Howard, 1960). In similar fashion, much of applied AI is focused on reinforcement learning (Kaelbling et al., 1996), which is also sequential, based on a Markov decision process (Agrawal, 2019). The sequential nature of applied decision making has also led to economists and AI scientists each modelling intelligent agents with Bayesian probability theory to inform their beliefs (priors), and utility theory to inform their preferences (Naudé, 2023).

This complex environment renders expected utility and behavioral economics unable to serve as core models in real-world applications, primarily due to their reliance on single-period lotteries with intertemporal choice augmentations such as exponential discounting (Samuelson, 1937) and quasi-hyperbolic discounting (O'Donoghue & Rabin, 1999), These approaches don't explicitly define the causal relationship that uncertain benefits follow upfront certain and uncertain costs over time on significant decisions (Horton, 2019). In contrast, the structure of causal economics (Horton, 2019) does align to this real-world environment, as this paper will explore. Causal economics provides a consistent theoretical underpinning to applied modelling in economics and AI that utilize dynamic stochastic programming and AI machine learning algorithms (Naudé, 2023).

The theoretical foundations of causal economics align strongly with the current surge of interest in causal inference throughout econometrics, statistics and computer science, represented in research on natural experiments (Imbens & Angrist, 1994) based on the potential outcomes approach (Rubin, 2005) of the Rubin causal model (Imbens & Rubin, 2010) and the recent development of causal machine learning (Huber, 2023) with its expanding set of literature (Arti et al., 2020). This invigorated and cross-disciplinary interest in causality presents an opportunity for causal economic theory (CE), causal machine learning (CML) and the natural experiments approach to be used together as a research agenda in causal economic machine learning (CEML).

2. Theory

2.1. Causal Economics Theory

There have been many significant advances in decision theory across academic disciplines, but more work needs to be done to provide a consistent, unified foundational framework, at the theoretical and empirical levels (Mishra, 2014). Causal economics attempts to address this gap by providing a robust framework well suited to testing a wide range of algorithms that address issues in AI research. The theory is grounded in behavioral economics, psychology, biology and neuroscience (Horton, 2019). In the following section, it will be demonstrated that expected utility (Mongin, 1997) and cumulative prospect theory (Tversky & Kahneman, 1992) also represent constrained mathematical reductions of causal economics that limit applications. The richness of causal economics is possible due to the definition of outcomes. Outcomes in causal economics must always contain both cost and benefit, where each of these contain certain (deliberate) and uncertain elements. Causation must also run in at least one direction of the relationship (Horton, 2019). The approach to preferences used in causal economics is built upon the structural model of cumulative prospect theory (Fennema & Wakker, 1997). With this foundation in place, preferences are expanded to include cost broadly (incorporating certain and uncertain components), including risk in the form of the uncertain component. Psychological trade-off-constraints are then defined as the primary umbrella constraint for optimization, which includes budget constraints (Horton, 2019). The

following summary briefly introduces the equations of the causal economic framework for those not familiar with it (Horton, 2019), beginning with key definitions:

i = an outcome where P incrementally results from Q and/or Q incrementally results from P (Q is required for P and/or visa versa)

t = time period

+,- = positive overall outcome and negative overall outcome respectively

(+) , (-) = personal total benefit and personal total cost outcome values respectively

w = probability weighting function

v = value function

π = capacity function of individual events

Ψ = rank dependent cumulative weighting of overall weighted event values

Ω = rank dependent cumulative probability weighting

P_i = perceived magnitude of positive element in outcome i

Q_i = perceived magnitude of negative element in outcome i

p_i = perceived probability of positive element in outcome i

q_i = perceived probability of negative element in outcome i

P^{C}_i = acceptable relative magnitude of positive element in outcome i

Q^{C}_i = acceptable relative magnitude of negative element in outcome i

p^{C}_i = acceptable relative probability of positive element in outcome i

q^{C}_i = acceptable relative probability of negative element in outcome i

$Ψ^{C}$ = acceptable rank dependent cumulative weighting of overall weighted event values

Z_u = perceived magnitude of outcome i (used when positive and negative outcomes aren't specified)

z_u = perceived probability of outcome i (used when positive and negative outcomes aren't specified)

Optimization in causal economics involves maximization of the value function, *v*, shown in equation (1);

$$
\begin{aligned}
Max \ v(f) = \int_{t=0}^{M} \Big\{ & \int_{i=0}^{N} [w^{+(+)}(p_{U\,i\,t})\,v^{+(+)}\,(P_{U\,i\,t}) + v^{+(+)}(P_{A\,i\,t}) \\
& - w^{+(-)}(q_{U\,i\,t})\,v^{+(-)}(Q_{U\,i\,t}) - v^{+(-)}(Q_{A\,i\,t})]\,\varphi_i^+ \\
& + \int_{i=-k}^{-1} [w^{-(+)}(p_{U\,i\,t})\,v^{-(+)}\,(P_{U\,i\,t}) + v^{-(+)}(P_{A\,i\,t}) \\
& - w^{-(-)}(q_{U\,i\,t})\,v^{-(-)}(Q_{U\,i\,t}) - v^{-(-)}\,(Q_{A\,i\,t})]\,\varphi_i^- \Big\}
\end{aligned}
$$

(1)

This maximization is subject to a pair of personal psychological trade-off constraints, defined by (2) and (3);

$$
c_i = \frac{\big[\,w^{(+)}(p^{C}_{U\,i\,t})v^{(+)}\,(P^{C}_{U\,i\,t}) + v^{+}(P^{C}_{A\,i\,t})\,\big]}{\big[\,w^{(-)}\,(q^{C}_{U\,i\,t})\,v^{(-)}(Q^{C}_{U\,i\,t}) + v^{(-)}(Q^{C}_{A\,i\,t})\,\big]} \times \varphi_i^{C}
$$

(2)

$$
L = \ Max\,[\,\{w^{(-)}(q^{C}_{U\,i\,t})\,v^{(-)}(Q^{C}_{U\,i\,t}) + v^{(-)}(Q^{C}_{A\,i\,t})\},\varphi_i^{C}\,]
$$

(3)

Optimization in causal economics differs fundamentally from the approach utilized in traditional economics, mainly through the addition of these two psychological trade-off constraints. Decision makers optimize by making trade-off decisions in light of their personal, internal psychological constraint of acceptable cost relative to causally coupled benefit (Horton, 2019). External constraints such as the decision maker's financial budget impact utility indirectly through the personal interpretation of the decision maker, based on that agent's personal psychological constraint, which is a matter of personal choice based on their individual experiences (Vahabi, 2001).

In causal economic optimization, the value function, *v*, and the psychological trade-off constraint, *c*, play separate but complementary roles. The value function captures a decision maker's personal judgements regarding the particular set of prospects they believe they may face when

making a decision and the specific costs and benefits they associate as realistically possible with each event. It incorporates preferences over both risk and time, but also represents perceived realities that the decision maker feels they must take as given (Horton, 2019). This is contrasted with the psychological trade-off constraint, *c*, that directly conveys a decision-maker's preferences across costs (uncertain and uncertain) as well as risk (uncertain). It conveys the marginal amount of benefit required for a decision maker to take on one additional unit of cost, or similarly the marginal amount of cost they will bear in order to obtain an additional unit of benefit.

The resulting optimization dynamic between v and c is therefore a process in which anticipated prospects are evaluated, via v, and compared to acceptable scenarios, via c, subject to L as a limit, at least in the short-term (Horton, 2019).

2.2. A Real World Example

A powerful way to illustrate how causal economics is able to effectively model the complex decisions agents face in the real-world, where traditional economic models cannot, is through a common example that most people can relate to—the pursuit of weight loss. A goal to lose weight requires the decision maker to weigh known costs and risks in advance of anticipated resulting future benefits (Horton, 2019). It is certainly more complex than going to the gym and eating a healthy diet for a day and then immediately losing a few pounds. Losing weight actually requires motivating goals (Brink & Ferguson, 1998) as well as deliberate and sustained commitment to incurring challenging personal costs in the form of dietary sacrifices and physical exertion that both extend over months (Horton, 2019). Neither the effort required nor the results obtained are primarily random lotteries, as would be utilized in traditional economic optimization (Horton, 2019).

Causal economics is able to effectively model this weight loss scenario because its framework builds in a multi-period structure, containing both certain (deliberate) and uncertain cost and benefit, that also requires costs to be causally matched to resulting benefits, both contemporaneously and in future tine periods. This causal coupling means that costs (some certain and some uncertain) are explicitly and directly tied to the expected resulting benefits that occur over the entire decision time-horizon (Horton, 2019). Costs are generally greater in earlier periods with relatively higher certainty as contrasted with potential benefits that can occur in later periods with relatively less certainty (Horton, 2019). In the weight loss example, costs that can be controlled to a large degree include exercise and diet. The associated causal gains involve weight loss, improved health and positive self-esteem.

The complexity of modelling an individual's decision to achieve a weight loss goal weight doesn't end with this matching of costs and benefits over the decision time-frame. Decision makers are each individuals with their own unique subjective preferences and subjective probability assessments, as captured by subjective expected utility theory (Savage, 1954). For example, an individual may pursue an exercise program in order to obtain their own perceptions of personal benefit, such as improved health and attractiveness (Brink & Ferguson, 1998), with full knowledge that exercising and eating well will allow them to achieve their goal with certainty, but only if they commit to the effort over time. There is uncertainty concerning the actual outcome quantities of individual parameters, but the decision maker's ability to deliberately decide and control the overall outcomes of their decisions (versus a lottery) is a large gap in expected utility and behavioral economics models (Horton, 2019). In the real world decision makers do not face a series of random lotteries. They impact and are impacted by outcomes (Horton, 2019). Many popular expressions in the English language capture this observed reality, such as 'getting out what you put in', 'no pain, no gain' and 'bearing the fruit of one's labour' (Horton, 2019).

2.3. Reducing Causal Economics to Expected Utility Theory

Expected utility (Schoemaker, 1982) can be directly derived from the model of causal economics with a number of restrictions. The first element of constraining causal economics to expected utility is restricting outcomes to single-value, net positive or negative lottery outcomes. This is accomplished by applying the following conditions to equations (1), (2) and (3):

$w^{+(+)}$, $w^{+(-)}$, $w^{-(+)}$, $w^{-(-)} = 0$

$v^{+(+)}$, $v^{+(-)}$, $v^{-(+)}$, $v^{-(-)} = 0$

p_A, q_A, P_A, $Q_A = 0$

p_U, q_U, P_U, $Q_U = 0$

w, v, Z, $z \neq 0$

Ψ, $\Psi^C = 1$

Given that outcomes in causal economics typically span a significant number of time periods, the second element of reducing causal economics to expected utility is the requirement that;

$$t = t_0$$

Based further on the assumption of linear independence as required for expected utility (Fishburn, 2013), the expected utility value function is defined by the familiar equation (4);

$$f = \sum_{i=0}^{N} [v(Z_{U\,i})\,(z_{U\,i})] \qquad (4)$$

In expected utility theory the value function is optimized in the context of standard consumer choice theory, based on indifference curves and budget constraints (Isaac, 1998). In order to reduce causal economic optimization to expected utility optimization, the internal psychological trade-off constraint function (2) must also be limited to a budget constraint and the personal total cost threshold constraint (3) must be allowed to take on an infinite value.

Time preferences are often added to the risk preferences of expected utility through the delta model (O'Donoghue & Rabin, 1999), which incorporates exponential discounting (Samuelson, 1937). Exponential discounting is time-separable, also known as time-consistent or dynamically consistent, which implies that a decision maker's trade-offs between utility today and delayed utility in the future, are independent of when that delay occurs. In the delta model, each subsequent period utility value is multiplied by δ^t as shown in equation (5);

$$f = \sum_{t=1}^{M} \delta^t \sum_{i=1}^{N} [U(Z_{U\,i})\,(z_{U\,i})] \qquad (5)$$

This approach to intertemporal time discounting applies the delta discount factor to the overall utility function, which is a reduction from the framework of causal economics, where time discounting can be captured within any or all of the weighting functions, $w^{+(+)}$, $w^{+(-)}$, $w^{-(+)}$, $w^{-(-)}$, $v^{+(+)}$, $v^{+(-)}$, $v^{-(+)}$, $v^{-(-)}$ with a functional form that fits the data appropriately (Horton, 2019).

2.4. Reducing Causal Economics to Behavioral Economics

Causal economics can also be mathematically reduced to cumulative prospect theory, which is the most robust formalized framework within behavioral economics. The first step in doing so is to restrict outcomes to single-value, net positive or negative lottery outcomes as was done with expected utility. This is accomplished by applying the following conditions to equations (1), (2) and (3);

$w^{+(+)}$, $w^{+(-)}$, $w^{-(+)}$, $w^{-(-)} = 0$

$v^{+(+)}$, $v^{+(-)}$, $v^{-(+)}$, $v^{-(-)} = 0$

p_A, q_A, P_A, $Q_A = 0$

p_U, q_U, P_U, $Q_U = 0$

w, v, Z, $z \neq 0$

Ψ, $\Psi^C = 1$

Outcomes in causal economics typically span a significant number of time periods, which results in the second requirement for reducing causal economics to cumulative prospect theory;

$$t = t_0 \qquad (5)$$

In cumulative prospect theory, Z and z are relative to status quo (instead of zero), Ω is a rank dependent cumulative probability weighting, and sign comonotonic trade-off consistency (Wakker

& Tversky, 1993) holds. Under all of these conditions, the causal economic value function (1) reduces to the utility value function for cumulative prospect theory represented by equation (6);

$$V(f) = \{\int_{i=0}^{N} \left[[v(Z_{U\,i})\,w(z_{U\,i})]\right]\Omega_i +$$
$$\int_{i=-k}^{-1} \left[[v(Z_{U\,i})\,w(z_{U\,i})]\,\right]\Omega_i\,\} \tag{6}$$

In cumulative prospect theory, the value function is optimized in the context of standard consumer choice theory, based on indifference curves and budget constraints (Isaac, 1998). As a result, in order to reduce causal economic optimization to cumulative prospect theory optimization, the internal psychological trade-off constraint function (2) must also be limited to a budget constraint and the personal total cost threshold constraint (3) must be allowed to take on an infinite value.

This reduction to cumulative prospect theory addresses the risk preference element of behavioral economics. However, intertemporal choice is an important component of behavioral economics (Dhami, 2016) and is typically incorporated through quasi-hyperbolic discounting via the beta-delta model (O'Donoghue & Rabin, 1999). Quasi-hyperbolic discounting captures time inconsistent preferences, such as present bias and changing preferences as time elapses. These are important additions, as extensive research demonstrates that decision makers generally do not have time consistent preferences (Thaler, 1981). They tend to be more patient for long-term gains and relatively impatient for short-term gains, which means that they often plan to do something in the future, but subsequently change their mind. These characteristics align closely with the concept of causal coupling, especially as illustrated in the real world example in section 2.2. Causal economics weighting functions often employ quasi-hyperbolic discounting (Horton, 2019).

In the beta-delta model, each subsequent period utility value is multiplied by β and δ^t as shown in equation (7).

$$V(f) = \beta\,\{\int_{i=0}^{N} \delta^t \left[[v(Z_i)\,w(p_i)]\right]\Omega_i +$$
$$\int_{i=-k}^{-1} \delta^t \left[[Z(Q_i)\,w(q_i)]\,\right]\Omega_i\,\} \tag{7}$$

This approach to intertemporal time discounting applies the beta and delta discount factors overall to the utility function, which is a reduction from the framework of causal economics, where time discounting can be captured within any or all of the weighting functions, $w^{+(+)}$, $w+^{(-)}$, $w^{-(+)}$, $w^{-(-)}$, $v^{+(+)}$, $v^{+(-)}$, $v^{-(+)}$, $v^{-(-)}$ with a functional form that fits the data appropriately (Horton, 2019).

2.5. Causal Economics and Artificial Intelligence

2.5.1. The Principle of Causal Coupling

2.5.1.1. Microeconomics

The previous sections have mapped out causal economics and demonstrated that it can be used as a single unified decision model that captures the breadth of scenarios applicable in the real world; those tackled by AI. This provides increased clarity for AI modelling at the individual level. AI is a broad discipline that attacks many problems with a range of techniques (Pannu, 2015).   However, the richness of causal economics creates complexity and there is a need to be aware of the risk of combinatorial explosion slowing down computation and impacting solvability (Russell & Norvig, 2021). Causal economics is most relevant to AI problems in the areas of planning, decision making and learning. These areas deal with intelligent reasoning in the face of uncertainty and incomplete information over time. As a result they incorporate probabilities, feedback and learning, utilizing various tools from probability and economics (Russel & Norvig, 2021). Research in adaptive economics has mapped out the importance of this feedback and learning in a complex and evolving environment (Day, 1975). The fundamental anchor in causal economics that aligns it closely with this context is the concept of causal coupling (CC). Causal coupling builds in cost and benefits that reflect

subjectivity, causality, and uncertainty and information availability over time. Causal coupling is formally defined in equation (8);

$$When \ldots \sum_{t=0}^{M} v^{(+)}\left(P_{Ait}\right) + v^{(+)}\left(P_{Uit}\right) > 0 \ldots \textit{ it holds } that \ldots$$

$$w^{(+)}\left(p_{Uit}\right) > 0, \sum_{t=0}^{M} v^{(-)}\left(Q_{Ait}\right) + v^{(-)}\left(Q_{Uit}\right) > 0 \ \& \ w^{(-)}\left(q_{Uit}\right) > 0 \ldots and \ldots \tag{8}$$

$$\sum_{t=0}^{M} t \, where \, \left\{v^{(+)}\left(P_{Ait}\right) + v^{(+)}\left(P_{Uit}\right)\right\} \ > \ 0 \ > \ \sum_{t=0}^{M} t \, where \, \left\{v^{(-)}\left(Q_{Ait}\right) + v^{(-)}\left(Q_{Uit}\right)\right\} > 0$$

Equation (8) requires that each outcome must contain both benefits and costs (contemporaneous and subsequent causation). It reflects the notion that decision makers in general have more immediate control over and face more immediate impact on their costs in terms of upfront effort than they do over the benefits that typically follow with some time delay. Equations (1), (2) and (8) directly build causal coupling into the definitions of percepts, outcomes, preferences and constraints. This mathematical grounding provides a robust foundation for AI algorithm development and application in individual agent decision making.

2.5.1.2. Macroeconomics

Whereas causal coupling connects cost and benefit for individual decision makers as a causal relationship over time, it can also be applied at a macro policy level in pursuit of societal optimization (Horton, 2019). This possibility centers on a bold conclusion of causal economics—that economic social coordination failures (outcomes that are not Pareto optimal) within societies are at their root the result of persistent causal decoupling of benefits and costs within populations of involved and/or impacted agents, and that effective solutions are the result of ensuring that causal coupling is achieved (Horton, 2019). In essence, social coordination failure is a generalization of the well-entrenched concept of market failure (Horton, 2019).

The pursuit of self-interest through the power of the free market, captured by Adam Smith's invisible hand metaphor has been demonstrated in practice (Baumol, 2002) but there are significant market failures in society that reduce welfare (Stiglitz, 1991). The application of causal coupling at a macro level attempts to close this gap through policies that couple cost and benefit for individuals, building in the notion of freedom with accountability (Horton, 2019). A loose metaphor for causal coupling at the macro level is perhaps the generalizing of Adam Smith's invisible hand into invisible hands. One hand represents the freedom to pursue self-interest for personal benefit, while the other represents the obligation to contribute a prorated share to societal costs as determined through democracy. This perspective allows a fresh opportunity for AI researchers to apply causal coupling as a macro policy condition when testing for policy optimality in populations.

Market failure occurs when the voluntary actions of agents fail to take into account broader externalities to the current transaction (Ledyard, 1989). Social coordination failure typically results when an agent is able to obtain personal enrichment without bearing the causally preceding costs, essentially gaining benefit through means other than a voluntary interaction/exchange, and/or avoiding their prorated share of the cost of public programs (Horton, 2019). It is essentially a causal decoupling of cost and benefit over the intermediate to long-term across individuals and society at large. In causal economic theory market failure is not due to exogenous externalities, but is instead the result of endogenous causal decoupling between cost and benefit across individuals and the population (Horton, 2019).

In practice, the largest source of causal decoupling in societies is when intermediaries receive/distribute compensation/benefits that is not tied to performance (Horton, 2019). Social policies are Pareto optimal when they couple benefits and costs for agents and do not allow intermediaries to accrue extra benefit from and/or offload extra cost to others that have no choice in the matter, taking into account realistic preferences, usage and risk (Horton, 2019). When involuntary allocations are applied through taxation to fund democratically determined public goods and

services and risk sharing programs, optimal allocations of benefit and cost are achieved by prorating costs across individuals either though a flat tax or user fee, with relief for those facing significant challenges to paying a proportionate share, subsidized by taxpayers with higher ability to pay (Horton, 2019). Causal economics asserts the optimality of taxation based on voter-approved specific usage with committed metrics (use-based taxes), as opposed to source-based taxation that is automatically collected on activities associated with economic value creation, such as income and consumption (Horton, 2019) and placed into general government coffers. This doesn't necessitate higher or lower taxes on individuals or overall. It just places the focus on a causally coupled, efficient means of raising funds for directed purposes desired by a democratic majority of voters.

2.6. The Neuroscience Underpinning Causal Economics

The concepts and structural setup of causal economics are closely aligned to findings in neuroscience (Kurniawan et al., 2010). Alongside psychology, neuroscience adds a fundamental perspective to our understanding of human decision making. This convergence in research provides a strong grounding for causal economics and its use as a foundation in AI. Studies in neuroscience have established the integration of reward and pain during choice (Talmi et al., 2009). It has been demonstrated that the cost required in action is a vital determinant of choice behavior (Kennerley et al., 2009). Multiple human and animal experiments have shown how the effort one must exert impacts choice and how the brain calculates and integrates effort into action value (Croxson et al., 2009; Floresco & Ghods-Sharifi, 2007; Kennerley et al., 2009; Rudebeck et al., 2008). Total effort is a combination of upfront and ongoing efforts (Horton, 2019) and can be expended as either mental work (Botvinick et al., 2009), or physical work (Kurniawan et al., 2010). Additional research in neuroscience reinforces the discounting of prospects where outcomes contain potential pain or loss (Seymour et al., 2007; Talmi et al., 2009) and are separated over time (Kable & Glimcher, 2008; McClure et al., 2007; Pine et al., 2009). Overall, these observations on effort, costs and outcome value over time from research in neuroscience ground the upfront and ongoing cost component of causal coupling at the centre of causal economics.

2.7. The Psychology Underpinning Causal Economics

The science of large and complex choices that take upfront and continued effort over time is comprehensively covered in the field of psychology (Newell et al., 2022). The causal economics framework essentially consolidates and provides a common formalized structure to many proven concepts in psychology (Horton, 2019). From a theoretical perspective, the structural design of causal economics relies heavily on many principles of psychology (Horton, 2019), as does that of behavioral economics (Camerer, 1999). This is an important foundation to a robust framework that can be most useful when used in AI. However, this alignment to psychology doesn't end with theory. In applied practice psychologists actively work to assist in the management of a decision maker's perceived costs and benefits to obtain desired outcomes. These therapies can focus on thoughts and behaviors to various degrees, through methods such as cognitive behavioral therapy, behavioral therapy or behavioral choice theory (Hollon, 2013).

2.8. Decision Making in Artificial Intelligence

AI seeks to address a broad range of challenges, spanning reasoning, planning and learning and perception, with applications that span areas such as software, robotics and the Internet of Things (Ertel, 2018). AI researchers have devised many tools to tackle these problems, and since AI agents generally operate with incomplete or uncertain information in the real world, for the most part methodologies from probability theory and economics are leveraged (Russell & Norvig, 2021). Modelling of decision making agents in AI typically parallels the approach of microeconomic decision making theory, based on bounded rational utility maximization in the face of constraints and perceptions of the environment (Naudé, 2023). Causal economics and its use of causal coupling provides the best structural fit as an underlying model of choice in the context of AI research.

AI takes the bold step of attempting to automate the decision process for efficiency and results that would parallel a human decision maker. Automated planning (Cimatti et al., 2008) involves agents pursuing a specific goal, in the face of complexity, where solutions need to be discovered and optimized in multidimensional space (Russell & Norvig, 2021). Decision makers choose an action with some known costs and make probabilistic guesses about future additional costs and benefits. In multi agent situations, the decision maker's environment and even preferences can be dynamically unknown, requiring ongoing iterative revision to the model and associated strategies and policies (Russell & Norvig, 2021). Agents are often forced to reassess their situation once results are experienced (Russell & Norvig, 2021). As the field of artificial intelligence is fundamentally concerned with solving applied problems, not just theory, a range of computational techniques are deployed, including dynamic programming, reinforcement learning and combinatorial optimization (Russell & Norvig, 2021). Game theory is utilized in situations where a researcher is modelling rational behavior across multiple interacting agents (Russell & Norvig, 2021).

Like applied economics, AI research typically utilizes dynamic stochastic programming (Bellman, 1957), based on a Markov decision process (Garcia et al., 2013). In this approach, a transition model is defined that describes the probability that a specific action will change the resulting state in a certain way as well as a reward function that defines the utility benefit value of each state and the cost of each action. Within this context, a policy is calculated or learned and associates a decision with each possible state of the world (Russell & Norvig, 2021).

There has been increasing momentum among AI researchers for the use of subjective probability, in order to model utility functions based on actual consequences and the decision taken, a priori probability distributions and other probability distributions impacting the decision (Suppes, 1969). Currently, AI researchers often model intelligent agents using Bayesian probability theory to inform beliefs (priors), and utility theory to inform preferences (Naudé, 2023). A broad range of practical problems are addressed with Bayesian tools, such as Bayesian networks (Domingos, 2015). Particular algorithms are often applied to particular classes of problems, such as the Bayesian inference algorithm for reasoning, the expectation-maximization algorithm for learning (Russell & Norvig, 2021), decision networks for planning using decision networks (Russell & Norvig, 2021), and dynamic Bayesian networks for perception (Poole et al., 1998).

It is clear that the environment in which AI research is conducted is one of complex decisions involving causally-linked cost and benefit trade-offs over extended periods of time, in the face of uncertainty, and across many agents. This environment is naturally modelled with causal economics and causal coupling, and does not correspond highly to the underlying frameworks of expected utility or behavioral economics (Horton, 2019).

2.9. Machine Learning

The previous section illustrated the underlying context in which AI agents must make decisions; one of subjective, causally-linked cost benefit trade-offs over time, in the face of uncertainty, with multiple interacting individuals and feedback from actions taken. But stopping at this point would forego the opportunity to incorporate learning insights that reinforce effective strategies while simultaneously avoiding ineffective ones (Barto et al., 1989). In real world decision making situations, humans iteratively discover information, without a priori knowledge of eventual results. They regularly experience outcomes that are demonstrated to be suboptimal posteriori, however these situations provide learning opportunities, and as a result decision makers are able to modify future decisions in pursuit of improved future results (Chialvo & Bak, 1999). The fundamental driver of the study of machine learning is to understand and apply programs that are able to improve their own performance on a particular task automatically (Russell & Norvig, 2021). The notion of computers being able to learn from their own decisions and outcomes has been fundamental to AI from the outset (Turing, 1950). Today, machine learning has become central to the resurgence of interest that is occurring in AI (Jayabharathi & Ilango, 2021). Fields such as healthcare are making very strong use of machine learning and seeing positive results (Bera et. al., 2023). The study of machine learning can fall into one of two major categories—unsupervised and supervised. In the unsupervised machine

learning approach, algorithms analyse data in order to find patterns and make predictions from that analysis automatically without any human guidance. In contrast, the supervised machine learning approach involves a human, with the role of labelling data inputs and categorizing it (Russell & Norvig, 2021).

Reinforcement learning is one of the most utilized forms of machine learning in practice and the approach aligns very closely with the principle of causal coupling. The approach of reinforcement learning applies rewards to the decision making agent for all actions that the model defines as good, and applies penalties to all actions that the model defines as bad (Luger & Stubblefield, 1990). Researchers have gone further, deriving the additional concept of inverse reinforcement learning. In such models decision making agents are able to seek out new information and then apply this experience-based knowledge to improve their preference functions. Inverse reinforcement learning does not utilize an explicit specification of a reward function because the reward function is actually inferred based on the observed behavior of the decision maker (Ng & Russell, 2000). The field of machine learning continues to evolve, bringing with it powerful new approaches that can improve the predictive power of models. An example is the inclusion of the concept of transfer learning, whereby knowledge obtained from a particular problem can be intelligently applied to a new problem to improve outcomes (Russel & Norvig, 2021). Building upon the previous discussion regarding the importance of grounding theory in multiple related disciplines, the area of deep learning is another core branch of machine learning that deserves attention. It leverages artificial neural networks that are inspired by biological systems with the objective of closely resembling the biological processes of human decision making (LeCun et. al., 2015).

The time-based, information-constrained and uncertain process of decision, action, feedback and learning captured in causal economics aligns directly with the context of machine learning. However, one component of causal economics that is missing in traditional machine learning methodology is causation, which is at the heart of causal machine learning and causal economics. Causal inference is addressed in the following section as a result.

2.10. Causal Inference

The previous section has sought to demonstrate that applied economics can certainly benefit from the use of machine learning in practice and that machine learning can benefit from explicit use of causal economics in its modelling. But the opportunity of these disciplines working in tandem for improved results has become greater than ever with the surge in research interest in two areas fundamentally focused on causality; natural experiments in econometrics and causal machine learning in computer science (Huber, 2023; Arti et al., 2020). Causality is at the very heart of causal economic theory through the concept of causal coupling (Horton, 2019), so it is particularly exciting that causality is seeing a surge of interest in the field of economics. This momentum has had an impact, as it is now considered a best practice in empirical work to identify all assumptions necessary to estimate causal effects (Lechner, 2023). Causal inference is vital in modern scientific research, because without it researchers only have correlations to rely on. In today's highly connect world, data almost always contains variables that impact each other, producing observations that are spurious and appear to have causal relationships that do not exist. It is clear that reliance on correlation cannot be relied upon for intervention decisions in areas such as healthcare (Sanchez et al., 2022).

2.10.1. Natural Experiments

A recent revolution in empirical research has brought the importance of causality front and center. The current prominence of causality in empirical work is highlighted clearly in the econometrics and statistics literature through the work of 2021 Nobel laureates David Card, Joshua Angrist and Guido Imbens on natural experiments (Hull et al., 2022) and independently in the biostatics literature through Baker and Lindeman (Baker & Lindeman, 1994). These insights build upon the potential outcomes framework (VanderWeele, 2016) first proposed by Jerzy Neyman and the subsequent Neyman–Rubin causal model (Imbens, 2015) devised by Donald Rubin. This invigorated and cross-disciplinary interest in causality presents an opportunity for causal economic

theory, causal machine learning and the natural experiments approach to be used together in future research.

Natural experiments are so powerful because they allow researchers to extract causal relationships from the real world, which is much more challenging to work with than data from a controlled clinical trial environment. In the latter, exposure to test and control groups can be controlled to ensure randomization. In natural experiments, individuals and/or groups are exposed to both the test and control conditions which are observed from nature rather than controlled by the experimenter. In nature there is often a confounding variable, one that is related to both the input variable and the output variable, which can distort the causal relationship of interest and even give rise to an apparent causal relationship that is actually spurious. Therefore, natural experiments attempt to approximate a random assignment. If a researcher suspects that changes in variable A cause changes in variable B, they compare the level of B across systems that vary in their level of A. So instead of manipulating A directly, the researcher analyzes the variations in the level of A observed in nature, not in a lab.

It is much more difficult to interpret cause and effect in natural experiments because individuals have themselves chosen whether they participate in the program/policy of interest. Even in this challenging context, robust conclusions can be drawn from natural experiments in which individuals cannot be forced or forbidden to participate in a policy/program (Angrist & Imbens, 1994), such that changes in outcomes may be plausibly attributed to exposure to the intervention (DiNardo, 2008). Since most randomised experiments in practice do not allow complete control over who actually participates in a policy intervention, these findings are broadly relevant and widely adopted by researchers in practice. Natural experiments can shed the most insight when there exists a clearly defined exposure to a policy/program intervention for a well-defined subpopulation and the absence of such exposure in a similar subpopulation. Though extremely powerful, the methods described can only provide an estimate of the effect of a policy intervention on the people who actually changed their behavior as a result of the natural experiment.

The Rubin causal model (RCM) measures the causal effect of a policy by comparing the two alternative potential outcomes that an individual could experience. Since only one can actually be observed as an outcome, one is always missing, an issue known as the fundamental problem of causal inference (Holland, 1986). The fundamental problem of causal inference means that causal effects cannot be directly measured at an individual level. A researcher must instead turn to population-level estimates of causal effects drawn from randomized experiments (Rubin, 1974). To the extent that an experiment can be randomized, individuals can be identified within one of two groups and then a difference in the means of the policy variable can be observed to ascertain the causal effect (Angrist & Imbens, 1995), known as the average treatment effect (ATE).

In natural experiments randomization is not always possible, as individuals may select particular outcomes based on other factors. In these cases, techniques such as propensity score matching (Kane et al., 2020) are utilized in an attempt to reduce treatment assignment bias, and mimic randomization. This is accomplished by creating a sample of individuals that received the treatment which is comparable on all observed covariates to a sample of units that did not receive the treatment. More precise estimates can be achieved by focusing only on individuals that are compliant (Hull, 2022), which leads to use of a measure such as the local average treatment effect (LATE), also known as known as the complier average causal effect (CACE) or the conditional average treatment effect (CATE). The LATE is typically calculated as either the ratio of the estimated intent-to-treat effect relative to estimated proportion of compliers, or through an instrumental variable estimator. However, this increase in precision comes with the trade-off of ignoring the effect of non-compliance to a policy that is typical in real world environments.

An alternative model of causal inference to the RCM is a structural causal model (SCM) (Pearl, 2010), which introduces a highly graphical approach which can be very helpful in modelling assumptions or identifying whether an intervention is even possible or not. The potential outcomes literature is more focused on quantifying the impact of policy/program interventions. In addition, there is ongoing work to potentially unify these methods, such as single world intervention graphs

(Sanchez et al., 2002). Continuing in the spirit of unification, there is further opportunity at present to explore whether predictive power in experiments can be increased by building the causal definitions of outcomes, preferences and constraints of causal economic theory into research models across broad areas such as computer science, economics, healthcare and others.

2.10.2. Causal Machine Learning (CML)

Causal machine learning (Huber, 2023) is currently an emerging area of computer science with an expanding literature (Arti & Kusumawardani, 2020). The increased focus on causality in empirical work spans many disciplines, such as computer science, econometrics, epidemiology, marketing, and statistics, as well as business (Lechner, 2023). Some of the most significant work underlying this area is the Neyman–Rubin causal model (Sekhon, 2007) based on the potential outcome approach of statistics (Imbens & Rubin, 2015) discussed in the previous section. Causal machine learning builds upon this underlying work and others in causal inference (Pearl, 2018) and applies machine learning methods (Lechner, 2023) to answer well identified causal questions using large and informative data (Athey, 2017). The continual pursuit of better treatment/policy outcomes by practitioners provides reason to believe that causal inference and machine learning (Cui et al., 2020) will come to play a dominant role in empirical analysis over time.

Pioneering economists are driving the increased use of causal machine learning in econometrics (Athey & Imbens, 2019) and its application to economic policy evaluation (Lechner, 2023). Prominent causal machine learning methods have been developed by economists directly, an example being causal forests as proposed by Susan Athey and Stefan Wager (Athey & Wager, 2019). The causal forest approach identifies neighbourhoods in the covariate space through recursive partitioning. Each causal tree within the causal forest learns a low-dimensional representation of the heterogeneity of the treatment effect. The causal forest is an average of a large number of individual causal trees, where each tree differs based on subsampling (Athey & Imbens, 2019). Studies find a value-added contribution from the use of causal machine learning in welfare economics applications (Strittmatter, 2023). With the continued expansion of the big data trend there will most likely be further increased interest in causal machine learning to deal with data sets that have a large numbers of covariates relative to the number of observations, posing a challenge to non-causal techniques.

Causal machine learning has experienced increasing popularity in healthcare (Sanchez et al., 2022) in tandem with a continued focus on real world evidence (RWE) and improvement of intervention outcomes (Crown, 2019). Obtaining valid statistical inference when using machine learning in causal research is currently a critical topic in public health and beyond (Naimi & Whitcomb, 2023). Medical researchers and practitioners are making some of the most powerful advances in the use of causal machine learning, asserting that machine learning promises to revolutionize clinical decision making and diagnosis (Richens et al., 2020). For example, a 2020 study conducted by Richens et al. showed that a causal, counterfactual algorithm delivered predictions that ranked within the top 25% of doctors, achieving expert clinical accuracy, compared to a standard associative algorithm that only placed in the 48% of doctors. These findings demonstrate that causal reasoning is a vital missing ingredient when applying machine learning to medical diagnosis situations (Richens et al., 2020). Recent research on the past expansion of health insurance programs using causal machine learning has provided tangible and robust insights into potential future programme expansions (Kreif et al., 2021).

A range of powerful meta-algorithms have been developed for causal machine learning, including the S-learner, the T-learner, the X-learner, the R-learner and the doubly robust (DR) learner. Each of these uses slightly different ways of estimating the average output and defining the conditional average treatment effect (Künzel, 2019). For example, in epidemiology, expert researchers suggest priority use of doubly robust estimators, the super learner and sample splitting to reduce bias and improve inference (Balzer & Westling, 2023). Work is ongoing to further categorize and apply causal machine learning (Kaddour et al., 2022).

The increased interest in causal machine learning is driven by the desire to use knowledge of a causal relationship to better predict outcomes of policies and treatments and to support confident

and informed decision making in important situations where reliance on spurious correlations could be dangerous or costly. Causal machine learning is where the theoretical and statistical approaches discussed previously see power in action in a world of big data. It is where the principles of causal economics connect directly with artificial intelligence, where theoretical and applied come together, unified in theory and practice.

2.10.3. Causal Economic Machine Learning (CEML)

Causal economic machine learning (CEML) represents the intersection of causal machine learning and causal economics. It essentially involves the replacement of traditional economic decision modelling with causal economic decision modelling in the practice of machine learning, particularly with respect to reinforcement learning and deep learning contexts. This primarily means observing outcomes that contain both certain and uncertain cost upfront that cause future uncertain benefits and costs when modelling individual decision makers. It can also mean application of causal coupling at the macroeconomic policy level to gauge the effectiveness of policies that do a better or worse job of coupling cost and benefit across individuals.

There is currently a tremendous amount of potential for a research agenda in CEML, as a causal machine learning framework that employs the model of causal economics—in particular studies that employ big data and the increasingly powerful software in use today.

3. Results and Discussion

3.1. Decision Making

The real world presents humans and machines with complex decision scenarios, involving causally-linked cost and benefit trade-offs over extended periods of time, in the face of uncertainty, while interacting with other competing and cooperating decision making agents. This context renders expected utility theory and behavioral economics theory unable to serve as core models in machine learning applications, primarily as a result of their reliance on single-period lotteries with intertemporal choice augmentations such as exponential discounting (Samuelson, 1937) and quasi-hyperbolic discounting (O'Donoghue & Rabin, 1999), that don't explicitly define the causal relationship that uncertain benefits follow upfront certain and uncertain costs over time on significant decisions (Horton, 2019).

In contrast, the framework of causal economics does provide an effective theoretical underpinning for use in AI and machine learning because it is structurally built to capture complex decision scenarios that involve causally-linked cost and benefit trade-offs over extended periods of time, in the face of uncertainty. It does not however explicitly build in multi-agent interaction dynamics, so additional methods such as game theory can be employed when defining an underlying model for an AI research program. Research has demonstrated the power of modelling inference with a combination of insight from game theory and machine learning (Rezek et al., 2008).

3.2. Macroeconomic / Social Outcomes

There is growing literature on the use of machine learning in macroeconomics (Hull & Hull, 2021), but it is hard to find much that utilizes causal machine learning in particular (Xu, 2018). Even without utilizing causal machine learning, the use of machine learning is producing improved results over full reliance on traditional econometric methods (Coulombe, 2020). Comprehensive research from the International Monetary Fund applies the elastic net, recurrent neural network (RNN) and super learner machine learning algorithms to broad economic analysis across the world and finds that overall they outperform traditional econometric techniques (Jung et al., 2018).

Causal economics and its core concept of causal coupling serve as a consistent tool for analyzing macroeconomic environments. It asserts broadly that structural institutions (ex. a free market, government body, central bank etc.) and intervention policies (ex. a tax policy etc.) will produce Pareto optimality when all decision makers and impacted stakeholders are able to enjoy the specific benefits that result from the costs they incur and bear the costs of the benefits they receive. A number

of characteristics reflect an economy with high levels of causal coupling in institutions and policies. A non-exhaustive list of a few prominent examples include the following (Horton, 2019) and provide for some novel experimental research:

- Free interaction and exchange (free markets) with corrections for externalities
- Significant levels of direct democracy
- Government by legislation prioritized over bureaucracy
- Widespread compensation for results in the private and public sector
- Taxation driven by flat and use based taxes with relief for those in need
- Social risk sharing programs for exposures that would be catastrophic to individuals

The application of causal coupling in macro policy would assert the optimality of conditions such as an engaged democratic majority, free competitive markets with corrections for externalities, a fair legal framework in place of government bureaucracies, minimal monetary and fiscal policy and taxes collected for specific voter-approved spending programs instead of being automatically collected and fed into general accounts based on income, consumption, property and wealth (Horton, 2019). Few of these conditions prevail together on a widespread scale in real economies, generally the result of vested interests and power, with a natural, biological motivation for those in power to pursue outcomes in their favour (Knauft, 1989).

3.2.1. Private Sector Application

Economic literature has a long history of demonstrating through precise models the power of competitive free markets to deliver Pareto optimal efficient allocations of goods and services through supply and demand (Baumol, 2002). Restrictions on competition, such as monopoly, duopoly, oligopoly and monopsony and government intervention through price and quantity controls are generally known to reduce this Pareto economic efficiency. Causal economics provides a simple and consolidated fresh new way to look at this. Competitive markets are optimal precisely because they maximize causal coupling of costs and benefits for those involved in the transaction (Horton, 2019). Buyers receive the perceived benefit of their purchase and part with the total personal cost (financial and/or psychological) of the item. Sellers part with the cost of the product/service they provide and obtain the benefit of funds that can be used for things they value more. Buyers and sellers make a voluntary decision on the trade-off between cost and benefit, which couples cost to benefit, and achieves Pareto optimality. It is clear that competitive free markets build causal coupling in directly at the transaction level (Horton, 2019).

However, transactions in the competitive free market are not overall optimal to the extent that they create involuntary externalities of costs and/or benefits to others. In these cases, government policies can correct the causal decoupling through tax/spend policies, for example through 'carbon taxes' to address environmental damage as directly as possible. The application of a tax/credit is essentially an effort to recouple cost and benefit. An example could involve a tax charged on producers of high pollution fuels that profit from the activity and the redistribution of those funds as a credit to companies and individuals that develop, deploy and utilize more environmentally sound, viable fuels.

Free and competitive markets are critical to freedom and economic prosperity, but they are not perfect. Private, profit-driven companies will understandably focus on segments of the market that are very willing and able to pay, often limiting availability and affordability of essential products and services, such as food and housing, to lower income individuals. Government intervention is not always at odds with the desires of taxpayers. There is much evidence that the majority of economic citizens support some degree of government-driven support to help those in serious need (Horton, 2019). Social safety nets exist in most advanced economies, demonstrating that even those that do not benefit directly from a program, but instead finance it, can still see value in its importance in society.

None of this understanding of markets and government intervention delivers new implications to existing economic theory; the purpose of this discussion is to illustrate how causal coupling provides a simple, consistent and universal way to look at economic and distributive societal issues that everyone can understand. At the level of the buyer and the firm, there is also an increased interest

in CML in digital marketing and ecommerce (Ma, 2020), where personalization is key to targeting profitable customers and particular customer engagement actions are optimal at each stage of the time-based, causal buying process. CEML provides an even great application fit in these areas than CML on its own, by incorporating the buyer's personal psychological trade-off preferences.

3.2.2. Public Sector Applications

The public sector plays an important role in most economics and has a large impact on the economy through taxation and government spending. As a result it has received a lot of attention in analysis. The use of deep reinforcement learning machine learning in analysis of taxation is providing quantifiable insights even without the explicit use of causality. An applied example is the empirical work done via Microsoft's The AI Economist project (Zheng et al., 2020). This study utilizes a two-level deep reinforcement learning approach in economic simulations to generate learning with respect to dynamic tax policies. In the model, both individuals and governments are able to learn and adapt (Zheng et al., 2020). The study demonstrates that AI-driven tax policies can improve the trade-off between equality and productivity by 16% over baseline policies.

Causal economic theory implies that taxation is optimal only when it is use-based, because this approach preserves the criterion of causal coupling (Horton, 2019). Use-based taxation requires that governments obtain advance approval from citizens prior to collection and are then subjected to compensation that is tied to performance goals. If a particular spending program is approved by voters then taxes are applied directly to cover the costs. This approach is in stark contrast to source-based taxation, which automatically collect taxes on economic variables such as income, consumption and wealth. Use-based taxation more closely parallels causally-coupled private sector funding, where entities must convince funding parties in advance that they will deliver results, based on their track record and plan.

Source-based taxation is non-optimal in causal economic theory because it directly decouples costs and benefits (Horton, 2019). The optimal alternatives that do achieve causal coupling are transparent flat taxes and user fees that are tied to results. Under such as system, tax relief is typically provided in democracies to those that are significantly challenged in contributing their prorated share. User fees are a very powerful causal-coupling mechanism because they directly couple the cost of the fee to the benefit the user obtains as directly as possible. Where specific users of particular government products/services can be identified, user fees are ideal. Where this is not the case, and it can be reasonably assumed that all citizens benefit from a particular government product/services, flat taxes are the ideal method of collection. Flat taxes spread the democratically-approved cost of a particular government program across all members of society, regardless of whether or not they personally benefit directly. In summary, causal coupling is maximized when government programs are funded through use-based tax policies, collect via user fees and/or flat taxes (Horton, 2019). Furthermore, the democratic majority can direct that tax relief and support be provided to individuals/groups that are unable to bear their prorated share of program costs.

There is an interesting opportunity to use CEML to test the effectiveness of changes in tax and/or spend programs as they represent clear policy interventions. Because source-based and use-based policies are so significantly contrasted, and because source-based taxes are extremely entrenched in economies across the world, there is exciting potential to use CEML test the effectiveness of incremental tax/spending programs with varying degrees of causal coupling.

3.2.3. Healthcare Applications

Healthcare often finds a home in the public and/or private sphere and is such a big impact area that it is considered here within its own category. Real-world evidence (RWE) has long been a prominent principle driving healthcare research (Khosla et al., 2018), and it remains a top priority due to its approach of evaluating patient interventions and outcomes based on data from actual clinical practice. By placing central focus on data-driven insight, RWE provides a natural point of extension into the application of machine learning into public health contexts (Kline et al., 2022; Li et al., 2020). The fit makes sense, because there has been a commensurate increase in the popularity of

CML in healthcare research (Sanchez et al., 2022). The most urgent challenges facing healthcare today require the comparison of the cost and benefit of alternative approaches, requiring that algorithms used can layout the consequences of a doctor's particular intervention treatment (Schulam & Saria, 2017). Promising new research continues to develop innovative CML methods that provide estimates of heterogeneous treatment effects in order to improve the individual-level, clinical decision-making of healthcare professionals in practice (Kuzmanovic, 2022).

Traditional machine learning is very effective in identifying risks and predicting outcomes, but it can't be relied upon when upon during clinical interventions, because in these settings, identification of causation is vital (Prosperi et al., 2020). Researchers have identified causal inference and reasoning as a critical yet often missing component in the application of machine learning to medical diagnosis (Richens et al., 2020). However, incorporating causation is not always easy in practice, since observational studies very often include confounding variables that impact each of the variables under consideration. The challenge of confounding biases in individualized medical interventions has been demonstrated through research that applies causal machine learning analysis to RWE data contained in electronic health records (EHR) (Zhang, 2023).

Additional research has shown the promising application of causal machine learning to evaluate understanding of the effectiveness of hospital discharges. Specifically, a study conducted at Kaiser Permanente Northern California (KPNC) healthcare demonstrated significant heterogeneous results in the responses of individual patients to intervention. In this study, causal machine learning was utilized in an attempt to identify preventable hospital readmissions, so they could be avoided. Results of this research illustrated that focusing on predicted treatment effect instead of predicted risk could potentially reduce readmissions by 30% annually (Marafino et al., 2020).

In each of the scenarios discussed, the introduction of causality into traditional correlation-based machine learning has been shown to provide much more realistic modelling of human decision making and behaviors through CML. CEML has extensive potential to further this progress by even more robustly modelling real human decision making and behaviors, including all of the benefits of CML and adding an additional layer of representation of the trade-offs between costs and benefits into the definitions of outcomes, preferences and constraints.

4. Conclusion

This paper has sought to demonstrate the unique and timely opportunity that currently exists to align the surge of interest in natural experiments and artificial intelligence, specifically causal machine learning, and tie them together with the framework of causal economics. There is a strong case for a specific research focus in causal economic machine learning (CEML). Beyond strengthened micro foundations, the implications of causal coupling provide an interesting approach in macro policy analysis and testing.

Researchers in AI are making rapid progress and the field is highly relevant and visible in society today. AI researchers focus on applied results, but are simultaneously interested in ensuring that their models are grounded in strong microeconomic decision modelling. Economic theory has made tremendous strides in many areas, but the dominant decision models of expected utility and behavioral economics have not provided a structure that works for AI and the real world focus of machine learning because they rely predominantly on lottery based, non-causal outcomes. Causal economics fills this gap through a framework that builds causality directly into the definitions of outcomes, preferences and trade-off constraints, incorporating multi-period decisions in the face of uncertainty, constrained information and limited computational capabilities. Additionally, causal economics leverages contributions to decision theory from many disciplines, including economics, psychology, neuroscience, statistics and computer science. To the extent that a single powerful model can be utilized, focus can be on solving real world problems, not a fragmented focus on building inconsistent models (Gerring, 2011).

Causal economic machine learning (CEML) sits at a convergence of three trajectories:
- *Expected utility theory and behavioral economics theory have led to causal economic theory.*

- *Causal inference (natural experiments, structural causal models) and machine learning have led to causal machine learning.*
- *Causal economic theory and causal machine learning have led to causal economic machine learning.*

There is exciting potential to adapt and develop algorithms that utilize the definitions of outcomes, preferences and trade-off constraints of causal economics which build in causality at the core. On this foundation there is also an opportunity to apply causal coupling as a constraint at the macro level to evaluate economic and social policy with causal economic machine learning.

5. References

1. Agrawal, S. (2019). Recent advances in multiarmed bandits for sequential decision making. Operations Research & Management Science in the Age of Analytics, 167-188.
2. Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of the American statistical Association, 431-442.
3. Arti, S., Hidayah, I., & Kusumawardani, S. S. (2020). Research trend of causal machine learning method: A literature review. IJID (International Journal on Informatics for Development), 9(2), 111-118.
4. Athey, S. (2018). The impact of machine learning on economics. In The economics of artificial intelligence: An agenda (pp. 507-547). University of Chicago Press.
5. Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. Annual Review of Economics, 11, 685-725.
6. Athey, S., & Wager, S. (2019). Estimating treatment effects with causal forests: An application. Observational studies, 5(2), 37-51
7. Baker, S. G., & Lindeman, K. S. (1994). The paired availability design: a proposal for evaluating epidural analgesia during labor. Statistics in Medicine, 13(21), 2269-2278.
8. Balzer, L. B., & Westling, T. (2023). Invited commentary: demystifying statistical inference when using machine learning in causal research. American Journal of Epidemiology, 192(9), 1545-1549.
9. Barto, A. G., Sutton, R. S., & Watkins, C. J. C. H. (1989). Learning and sequential decision making (pp. 89-95). Amherst, MA: University of Massachusetts.
10. Baumol, W. J. (2002). The free-market innovation machine: Analyzing the growth miracle of capitalism. Princeton university press.
11. Bera, S., Bali, S. K., & Kaur, R. (2023, June). Resurgence of artificial intelligence in healthcare: A survey. In AIP Conference Proceedings (Vol. 2705, No. 1). AIP Publishing.
12. Bellman, R. (1957). A Markovian decision process. Journal of mathematics and mechanics, 679-684.
13. Botvinick, M. M., Huffstetler, S., & McGuire, J. T. (2009). Effort discounting in human nucleus accumbens. Cognitive, affective, & behavioral neuroscience, 9, 16-27.
14. Brink, P. J., & Ferguson, K. (1998). The decision to lose weight. Western Journal of Nursing Research, 20(1), 84-102.
15. Camerer, C. (1999). Behavioral economics: Reunifying psychology and economics. Proceedings of the National Academy of Sciences, 96(19), 10575-10577.
16. Cimatti, A., Pistore, M., & Traverso, P. (2008). Automated planning. Foundations of Artificial Intelligence, 3, 841-867.
17. Croxson, P. L., Walton, M. E., O'Reilly, J. X., Behrens, T. E., & Rushworth, M. F. (2009). Effort-based cost–benefit valuation and the human brain. Journal of Neuroscience, 29(14), 4531-4541.
18. Coulombe, Philippe Goulet. How is Machine Learning Useful for Macroeconomic Forecasting?∗ Philippe Goulet Coulombe1† Maxime Leroux2 Dalibor Stevanovic2‡ Stéphane Surprenant2 1University of Pennsylvania 2Université du Québec à Montréal First version: October 2019 This version: August 24, 2020
19. Cui, P., Shen, Z., Li, S., Yao, L., Li, Y., Chu, Z., & Gao, J. (2020, August). Causal inference meets machine learning. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 3527-3528).
20. Chaturvedi D, Patvardhan C and Lakshmi, C. V. "AI Value Alignment Problem: The Clear and Present Danger," 2023 6th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2023, pp. 1-6, doi: 10.1109/ISCON57294.2023.10112100.
21. Chialvo, D. R., & Bak, P. (1999). Learning from mistakes. Neuroscience, 90(4), 1137-1148.
22. Crown, W. H. (2019). Real-world evidence, causal inference, and machine learning. Value in Health, 22(5), 587-592.19
23. Day, R. H. (1975). Adaptive processes and economic theory. In Adaptive economic models (pp. 1-38). Academic Press.
24. Dhami, S. (2016). The foundations of behavioral economic analysis. Oxford University Press.
25. Dillon, S. M. (1998, August). Descriptive decision making: Comparing theory with practice. In Proceedings of 33rd ORSNZ Conference, University of Auckland, New Zealand.

26. DiNardo, J. (2008). "Natural experiments and quasi-natural experiments". In Durlauf, Steven N.; Blume, Lawrence E (eds.). The New Palgrave Dictionary of Economics (Second ed.). Palgrave Macmillan. pp. 856–864. doi:10.1057/9780230226203.1162. ISBN 978-0-333-78676-5.

27. Domingos, P. (2015). The master algorithm: How the quest for the ultimate learning machine will remake our world. Basic Books.

28. Ertel, W. (2018). Introduction to artificial intelligence. Springer.

29.

30. Fennema, H., & Wakker, P. (1997). Original and Cumulative Prospect Theory: A discussion of empirical differences. Journal of Behavioral Decision Making, 10(1), 53-64.

31. Fishburn, P. C. (2013). The foundations of Expected Utility (Vol. 31). Springer Science & Business Media.

32. Floresco, S. B., & Ghods-Sharifi, S. (2007). Amygdala-prefrontal cortical circuitry regulates effort-based decision making. Cerebral cortex, 17(2), 251-260.

33. Gabriel, I., & Ghazavi, V. (2021). The challenge of value alignment: From fairer algorithms to AI safety. arXiv preprint arXiv:2101.06060.

34. Garcia, F., & Rachelson, E. (2013). Markov decision processes. Markov Decision Processes in Artificial Intelligence, 1-38.

35. Gerring, J. (2011). Social science methodology: A unified framework. Cambridge University Press.

36. Green, S. L. (2002). Rational choice theory: An overview. In Baylor University Faculty development seminar on rational choice theory (pp. 1-72).

37.

38. Holland, Paul W. (1986). "Statistics and Causal Inference". J. Amer. Statist. Assoc. 81 (396): 945–960. doi:10.1080/01621459.1986.10478354. JSTOR 2289064.

39. Hollon, S. D., & Beck, A. T. (2013). Cognitive and cognitive-behavioral therapies. Bergin and Garfield's handbook of psychotherapy and behavior change, 6, 393-442.

40. Horton, A. (2019). Causal Economics: A new pluralist framework for Behavioral Economics that advances theoretical and applied foundations. Heliyon, 5(4), e01342.

41. Howard, R. A. (1960). Dynamic programming and markov processes.

42. Huber, M. (2023). Causal analysis: Impact evaluation and Causal Machine Learning with applications in R. MIT Press

43. Hull, I., & Hull, I. (2021). Machine Learning and Economics. Machine Learning for Economics and Finance in TensorFlow 2: Deep Learning Models for Research and Industry, 61-86.

44. Hull, P., Kolesár, M., & Walters, C. (2022). Labor by design: contributions of David Card, Joshua Angrist, and Guido Imbens. The Scandinavian Journal of Economics, 124(3), 603-645.

45. Imbens, G. W., & Angrist, J. D. (1994). of local average treatment Journal of the Econometric. Identification and estimation effects. Econometrica: Society, 5, 467-475.

46. Imbens G, Rubin D. (2015). Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction. Cambridge University Press; 2015. doi:10.1017/CBO9781139025751

47. Imbens, G. W., & Rubin, D. B. (2010). Rubin causal model. In Microeconometrics (pp. 229-241). London: Palgrave Macmillan UK.

48. Isaac, A. G. (1998). The structure of neoclassical consumer theory (No. 9805003). University Library of Munich, Germany.

49. Jayabharathi, S., & Ilango, V. (2021, August). A Brief Revolution of Evolution and Resurgence on Machine Learning. In 2021 Asian Conference on Innovation in Technology (ASIANCON) (pp. 1-5). IEEE.

50. Jenkins, P., Farag, A., Jenkins, J. S., Yao, H., Wang, S., & Li, Z. (2021, May). Neural utility functions. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 9, pp. 7917-7925).

51. Jung, J. K., Patnam, M., & Ter-Martirosyan, A. (2018). An algorithmic crystal ball: Forecasts-based on machine learning. International Monetary Fund.

52. Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. Neuron, 63(6), 733-745.

53. Kaddour, J., Lynch, A., Liu, Q., Kusner, M. J., & Silva, R. (2022). Causal machine learning: A survey and open problems. arXiv preprint arXiv:2206.15475.

54. Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. Journal of artificial intelligence research, 4, 237-285.

55. Kane, L. T., Fang, T., Galetta, M. S., Goyal, D. K., Nicholson, K. J., Kepler, C. K., ... & Schroeder, G. D. (2020). Propensity score matching: a statistical method. Clinical spine surgery, 33(3), 120-122.

56. Kennerley, S. W., Dahmubed, A. F., Lara, A. H., & Wallis, J. D. (2009). Neurons in the frontal lobe encode the value of multiple decision variables. Journal of cognitive neuroscience, 21(6), 1162-1178.

57. Khosla, S., White, R., Medina, J., Ouwens, M., Emmas, C., Koder, T., ... & Leonard, S. (2018). Real world evidence (RWE)–a disruptive innovation or the quiet evolution of medical evidence generation?. F1000Research, 7.

58. Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., ... & Luo, Y. (2022). Multimodal machine learning in precision health: A scoping review. npj Digital Medicine, 5(1), 171.

59. Knauft, B. M. (1989). Sociality versus self-interest in human evolution. Behavioral and Brain Sciences, 12(4), 712-713.

60. Kreif, N., DiazOrdaz, K., Moreno-Serra, R., Mirelman, A., Hidayat, T., & Suhrcke, M. (2021). Estimating heterogeneous policy impacts using causal machine learning: a case study of health insurance reform in Indonesia. Health Services and Outcomes Research Methodology, 1-36.

61. Künzel, S. R., et. al. (2019). Metalearners for estimating heterogeneous treatment effects using machine learning.

62. Kurniawan, I. T., Seymour, B., Talmi, D., Yoshida, W., Chater, N., & Dolan, R. J. (2010). Choosing to make an effort: the role of striatum in signaling physical effort of a chosen action. Journal of neurophysiology, 104(1), 313-321.

63. Kuzmanovic, M. (2022). Advances in Causal Machine Learning for Health Management (Doctoral dissertation, ETH Zurich).

64. Lechner, M. Causal Machine Learning and its use for public policy. Swiss J Economics Statistics 159, 8 (2023). https://doi.org/10.1186/s41937-023-00113-y

65. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

66. Ledyard, J. O. (1989). Market failure. In Allocation, Information and Markets (pp. 185-190). London: Palgrave Macmillan UK.

67. Li, R., Yin, C., Yang, S., Qian, B., & Zhang, P. (2020). Marrying medical domain knowledge with deep learning on electronic health records: a deep visual analytics approach. Journal of medical Internet research, 22(9), e20645.

68. Luger, G. F., & Stubblefield, W. A. (1990). Artificial intelligence and the design of expert systems. Benjamin-Cummings Publishing Co., Inc..

69. Ma, L., & Sun, B. (2020). Machine learning and AI in marketing–Connecting computing power to human insights. International Journal of Research in Marketing, 37(3), 481-504.

70. Marafino, B. J., Schuler, A., Liu, V. X., Escobar, G. J., & Baiocchi, M. (2020). Predicting preventable hospital readmissions with causal machine learning. Health Services Research, 55(6), 993-1002.

71. McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. Journal of neuroscience, 27(21), 5796-5804.

72. Mishra, S. (2014). Decision-making under risk: Integrating perspectives from biology, economics, and psychology. Personality and Social Psychology Review, 18(3), 280-307.

73. Mongin, P. (1997). Expected Utility Theory. Handbook of economic methodology, 342350.

74. Naimi, A. I., & Whitcomb, B. W. (2023). Defining and Identifying Average Treatment Effects. American Journal of Epidemiology, 192(5), 685-687.

75. Naudé, W. (2023). Artificial Intelligence and the Economics of Decision-Making.

76. Newell, B. R., Lagnado, D. A., & Shanks, D. R. (2022). Straight choices: The psychology of decision making. Psychology Press.

77. Ng, A. Y.; Russell, S. J. (2000). "Algorithms for Inverse Reinforcement Learning" (PDF). Proceeding ICML '00 Proceedings of the Seventeenth International Conference on Machine Learning. pp. 663–670. ISBN 1-55860-707-2.

78. O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. American economic review, 89(1), 103-124.

79. Pannu, A. (2015). Artificial intelligence and its application in different areas. Artificial Intelligence, 4(10), 79-84.

80. Pearl, J. (2010). Causal inference. Causality: objectives and assessment, 39-58.

81. Pearl, J (2018). The Seven Tools of Causal Inference with Reflections on Machine Learning.

82. Pine, A., Seymour, B., Roiser, J. P., Bossaerts, P., Friston, K. J., Curran, H. V., & Dolan, R. J. (2009). Encoding of marginal utility across time in the human brain. Journal of Neuroscience, 29(30), 9575-9581.

83. Poole, D. I., Goebel, R. G., & Mackworth, A. K. (1998). Computational intelligence (Vol. 1). New York: Oxford University Press.

84. Prosperi, M., Guo, Y., Sperrin, M., Koopman, J. S., Min, J. S., He, X., ... & Bian, J. (2020). Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nature Machine Intelligence, 2(7), 369-375.

85. Rezek, I., Leslie, D. S., Reece, S., Roberts, S. J., Rogers, A., Dash, R. K., & Jennings, N. R. (2008). On similarities between inference in game theory and machine learning. Journal of Artificial Intelligence Research, 33, 259-283.

86. Richens, J. G., Lee, C. M., & Johri, S. (2020). Improving the accuracy of medical diagnosis with causal machine learning. Nature communications, 11(1), 3923.

87. Rubin, Donald (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies". J. Educ. Psychol. 66 (5): 688–701 [p. 689]. doi:10.1037/h0037350.

88. Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. Journal of the American Statistical Association, 100(469), 322-331.

89. Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., & Rushworth, M. F. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. Journal of Neuroscience, 28(51), 13775-13785.

90. Russell, Stuart J.; Norvig, Peter. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Hoboken: Pearson. ISBN 978-0134610993. LCCN 20190474.

91. Samuelson, P. A. (1937). A note on measurement of utility. The review of economic studies, 4(2), 155-161.

92. Sanchez Pedro,Voisey Jeremy P.,Xia Tian,Watson Hannah I.,O'Neil Alison Q. and Tsaftaris Sotirios A.2022 Causal machine learning for healthcare and precision medicineR. Soc. open sci.9220638220638

93. Savage, Leonard J. 1954. The Foundations of Statistics. New York, Wiley.

94. Schoemaker, P. J. (1982). The expected utility model: Its variants, purposes, evidence and limitations. Journal of economic literature, 529-563.

95. Schulam, P., & Saria, S. (2017). Reliable decision support using counterfactual models. Advances in neural information processing systems, 30.

96. Sejnowski, T. J. (2018). The deep learning revolution. MIT press.

97. Sekhon, Jasjeet (2007). "The Neyman–Rubin Model of Causal Inference and Estimation via Matching Methods" (PDF). The Oxford Handbook of Political Methodology.

98. Seymour, B., Daw, N., Dayan, P., Singer, T., & Dolan, R. (2007). Differential encoding of losses and gains in the human striatum. Journal of Neuroscience, 27(18), 4826-4831.

99. Simon, H. A. (1990). Bounded rationality. Utility and probability, 15-18.

100. Şimşek, Ö. (2020). Bounded rationality for artificial intelligence. In Routledge handbook of bounded rationality (pp. 338-348). Routledge.

101. Stiglitz, J. E. (1991). The invisible hand and modern welfare economics.

102. Strittmatter, A. (2023). What is the value added by using causal machine learning methods in a welfare experiment evaluation?. Labour Economics, 84, 102412.

103. Suppes, P. (1969). The role of subjective probability and utility in decision-making. In Studies in the Methodology and Foundations of Science: Selected Papers from 1951 to 1969 (pp. 87-104). Dordrecht: Springer Netherlands.

104. Talmi, D., Dayan, P., Kiebel, S. J., Frith, C. D., & Dolan, R. J. (2009). How humans integrate the prospects of pain and reward during choice. Journal of Neuroscience, 29(46), 14617-14626.

105. Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. Economics letters, 8(3), 201-207.

106. Turing, A. M. (1950). Mind. Mind, 59(236), 433-460.

107. Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. Journal of Risk and uncertainty, 5(4), 297-323.

108. Tversky, A. (1975). A critique of expected utility theory: Descriptive and normative considerations. Erkenntnis, 163-173.

109. Vahabi, M. (2001). The soft budget constraint: a theoretical clarification. Recherches Économiques de Louvain/Louvain Economic Review, 67(2), 157-195.

110. VanderWeele T. J. (2016). Commentary: On Causes, Causal Inference, and Potential Outcomes. International journal of epidemiology, 45(6), 1809–1816. https://doi.org/10.1093/ije/dyw230

111. Wakker, P., & Tversky, A. (1993). An axiomatization of Cumulative Prospect Theory. Journal of risk and uncertainty, 7(2), 147-175. https://doi.org/10.1007/bf01065812

112. Xu, L. (2018, December). Machine learning and causal analyses for modeling financial and economic data. In Applied Informatics (Vol. 5, No. 1, pp. 1-42). SpringerOpen.

113. Yaqub, M. Z., Saz, G., & Hussain, D. (2009). A meta analysis of the empirical evidence on expected utility theory. European Journal of Economics, Finance, and Administrative Sciences, 15, 117-133.

114. Zhang, L. (2023). Causal Machine Learning for Reliable Real-World Evidence Generation in Healthcare (Doctoral dissertation, Columbia University).

115. Zheng, S., Trott, A., Srinivasa, S., Naik, N., Gruesbeck, M., Parkes, D. C., & Socher, R. (2020). The ai economist: Improving equality and productivity with ai-driven tax policies. arXiv preprint arXiv:2004.13332.