Repository Entry Template
Embedded EthiCS @ Harvard Teaching Lab

Overview

Course: CS 236R: Topics at the Interface between Computer Science and Economics: Behavioral Economics and Computation

Course Level: Graduate

Course Description: "This is a rotating topics course that studies the interplay between computation and economics. The class is seminar style and readings are drawn from artificial intelligence, theoretical computer science, machine learning, multi-agent systems, economics, psychology and operations research.

The topic of Spring 2020 is behavioral economics and computation. The interdisciplinary field of Economics and Computation (EC) often takes a design perspective, attempting to develop systems (or mechanisms) that achieve certain system-wide goals while taking into consideration the behavior of their human participants. The rational agent model is widely adopted in this field to capture the human behavior in such systems. However, abundant evidence from psychology and behavioral economics has shown that human behavior deviates from the simple rational agent model. Since a theory is only as good as its model is, how can we integrate more realistic behavioral models in understanding and designing systems? How can we design better systems when we do not fully understand human behavior? Can a data-driven, computational approach help? What are some ethical considerations for designing such systems? We hope to expose students to a diverse set of emerging topics at the intersection of behavioral sciences and computer science."

Module Topic: Ethics of Digital Nudging

Module Author: Meica Magnani

Semesters Taught: Fall 2020

Tags: interface design [CS], big data [CS], nudging [both], hyper-nudging [both], transparency [both], direct and indirect stakeholders [phil], autonomy [phil], manipulation [phil], exploitation [phil], paternalism [phil]

Module Overview: In this module we discuss the ethics of the *nudge*. A nudge is an alteration in the decision environment that aims to influence human behavior without restricting options or introducing economic incentives. It is both a central concept in behavioral economics and a very common design-based mechanism implemented by tech companies to guide the attention, decisions, and behavior of users.

Students are introduced to the nudge, how it works, and its exciting potential for influencing human behavior and decisions. We consider how it can help users navigate the digital space and how big data is being used to create highly personalized and dynamic nudges (the *hypernudge*). We then take a look at some cases of nudging and hypernudging that strike many as morally problematic and try to identify what exactly, if anything, grounds the concern. Issues of paternalism, autonomy, manipulation, and

	exploitation are discussed. Students are also shown how to identify different stakeholders. The module ends with an activity in which students use stakeholder analysis and the ethical concepts discussed to propose alterations to the design of Uber/Lyft driver apps. They are asked to justify and motivate their design choices using the ethical concepts discussed.
Connection to Course Material:	Two of the central questions in the course are "how can we integrate more realistic behavioral models in understanding and designing systems?" and "can a data-driven, computational approach help [us model actual human behavior]?" The nudge is a design-based mechanism based on a more realistic understanding of human behavior, namely one that recognizes our fallibility, weakness of will, and sensitivity to environmental cues. Nudges are designed in light of these cognitive and motivational weaknesses. They can help us override our shortcomings. There are, however, many ethical questions that arise when we intentionally target the cognitive and motivational shortcomings of persons to influence their behavior. Using a computational approach to further identify these limitations of our psychology raises another dimension of concern.

	Goals
Module Goals:	1. Introduce students to nudges and hypernudges. Explain how they work. 2. Show students the potential of using nudges and hypernudges in design. 3. Equip students with tools for thinking through the ethics of nudges and hypernudges. 4. Give students practice using these tools to: (a) diagnose particular nudges (identify the ethical considerations and concerns); (b) design more ethical nudges and hypernudges; and (c) defend their design choices.
Key Philosophical Questions:	1. How might a nudge be paternalistic? Why might this be a concern? Why might it be desirable? 2. When and how could a nudge respect the autonomy of a user? When and how could it violate the autonomy of a user? 3. How might a nudge be manipulative? Are there ways to make nudges less manipulative? 4. How might a nudge be exploitative? What needs to be in place to make a nudge less exploitative?

	5. Who are the stakeholders for a given nudge? What are their interests? 6. How might the transparency of nudges or the consent to being nudged protect or promote autonomy?

Materials

Key Philosophical Concepts:	● Paternalism ● Autonomy ● Manipulation ● Exploitation ● Indirect and direct stakeholders ● Consent ● Transparency	Philosophers debate the precise definitions of these concepts. The Embedded EthiCS TA uses the following definitions: Paternalism: the making of decisions for another person which are in their supposed best interest. Autonomy: the capacity to think and choose for oneself. Manipulation: a kind of influence that does not sufficiently engage the capacities for reflective and deliberative choice of the person being influenced. Exploitation: when someone in a position of power takes advantage of a vulnerability in less powerful positions, in order to advance their own interests. One might also include Sunstein and Thaler's (the authors of nudge theory) concept of "Libertarian Paternalism" as a guide for permissible nudges. Unfortunately, they offer different definitions throughout their writings. According to Libertarian Paternalism, a nudge must: (1) respect the freedom of choice of individuals; and (2) be in the best interest of those nudged. (They are inconsistent on what "best interest" means: sometimes they say it is according to some objective measure other times they say "as judged by themselves.")

Assigned Readings:	● "Uber Shows How Not to Apply Behavioral Economics" *Harvard Business Review*	This reading gives a very basic explanation of nudges and points out how they can be used for good. The author then explains in simple terms (misalignment of interests) why the Uber app fails to be a good nudge.

	Implementation	
Class Agenda:	1. What is a nudge? A hypernudge? How do they work? 2. The potential of nudges. 3. Ethical concerns and problematic nudges. 4. Stakeholder Analysis: whose interests are at stake? 5. Class Activity: Uber/Lyft apps. Use ethics to diagnose the problems and propose a new design feature.	When discussing the potential of nudges, the Embedded EthiCS TA should bring out cases where nudges really do help users navigate digital spaces given their cognitive and motivational limitations (e.g. setting the strictest privacy setting for apps as the default option, Google search as a way to effectively wade through the sea of information on the internet, etc.). When discussing potentially problematic nudges, the idea is to get the students to identify the ethical concerns on their own. Ideally, the Embedded EthiCS TA is able to guide the discussion so as to extract and distill concerns about paternalism, autonomy, manipulation, and exploitation from the students. The TA should also draw out how lack of consent, misalignment of ends, and lack of transparency are various ways in which nudges can fail to respect autonomy, be manipulative, or be exploitative. Once these tools are on the table, the Embedded EthiCS TA contextualizes them within a stakeholder analysis of one of the dark nudges. Direct stakeholders: Uber/Lyft, drivers. Indirect stakeholders: passengers, businesses like bars that depend on transportation services, etc.

		Both the diagnosis and design discussion of the Uber/Lyft apps are to take place in small groups. Students discuss and brainstorm together. They then explain their diagnosis and propose their design ideas to the class. They are asked to defend their ideas in conversation with the class.
Sample Class Activity:	**Diagnose**. Carefully consider the controversial Uber/Lyft app. 1. Identify direct stakeholders and stakeholder interests. Are their interests aligned? 2. Identify indirect stakeholders and indirect stakeholder interests. Are their interests aligned with the direct stakeholders? 3. Does the nudge respect the autonomy of the nudged? Is it manipulative of the nudged? Is it exploitative of the nudge? WHY? Demonstrate understanding of concepts in your explanations. **Design.** Focus on one concern (e.g. if the autonomy of the driver is being compromised, if the nature of the nudge is manipulative, etc.). Propose a design feature change that eliminates or reduces this worry. Then explain how it impacts the interests of the other stakeholders (direct and indirect). .	
Module Assignment:	Students are given the following prompt: You work at Facebook. Facebook is very good at predicting when people are going through the emotional aftermath of a break-up. 1. **Design**: a nudge, either in the interest of Facebook (e.g. serving ads at the right time to the right people, collecting more users, etc.) or in the interest of the broken-hearted (e.g. emotional healing). Explain how the nudge works and serves the intended interest. 2. **Diagnose**: Who are the stakeholders? What are their interests? What are the different ethical concerns that someone might have with this nudge (even if it is a well-intentioned nudge!)? Be sure to thoroughly explain these concerns. 3. **Defend**. Ultimately, do you think it is okay to use this nudge? Why or why not? Use	This assignment reinforces concepts and skills learned in the module. Students first draw from their knowledge of nudges to design a nudge for Facebook. They then implement the normative concepts and philosophical tools to provide an ethical analysis of the nudge. They then are asked to think through whether or not it would be ethical for Facebook to use the nudge and to defend their ultimate position (drawing from ethical considerations learned in class).

	some of the ethical concepts learned in class.	
Lessons Learned Students find the concept of the nudge very interesting.The module benefits greatly from having lots of concrete examples to illustrate nudges, hypernudges, and dark nudges.		