4th IFAC Workshop on Distributed Estimation
and Control in Networked Systems
September 25-26, 2013. Rhine-Moselle-Hall, Koblenz, Germany

On informational nudging and control of payoff-based learning

Robin Guers [*] **Cedric Langbort** [**] **Dan Work** [***]

[*] *Aerospace Engineering, UIUC, Urbana, USA, (e-mail: guers2@illinois.edu).*
[**] *Aerospace Engineering & CSL, UIUC, Urbana, IL 61820, USA, (e-mail: langbort@illinois.edu).*
[***] *Civil & Environmental Engineering & CSL, UIUC, Urbana, IL 61820, USA, (e-mail: dbwork@illinois.edu).*

Abstract: We investigate a model of *informational nudging* in a context inspired by repeated games in traffic. Starting from a simple payoff–based learning model for an individual *decision–maker* (DM) choosing among multiple alternatives, we introduce a *recommender* who provides possibly misleading payoff information for unchosen options, so as to drive the DM's preferences to a desired equilibrium. This kind of white lie on the part of the recommender can be seen as an informational nudge in the sense of Thaler & Sunstein, and may thus arguably present some benefits over monetary incentive–based strategies for the purposes of planning. Considering the fluid limit of our simplified model, we show that the recommender can create (but not necessarily globally stabilize) any outcome he desires using constant lying strategies. We also identify a framing effect, in the sense that lies about the least favorable option has a different effect compared to lies on most favorable option.

Keywords: traffic control, stochastic approximation, learning

1. INTRODUCTION

One of the current challenges in the design and operation of large–scale smart infrastructures is the synergetic integration of cyber, human, and physical components, in other words the principled engineering of *cyber–socio–physical* systems. For example, existing proposals for the next generation of navigation and congestion management systems not only call for better traffic estimation and sensing algorithms, software, and hardware, but also for reliable mechanisms to influence commuters to modify their route and transportation mode choices.

Such mechanisms continue to be actively explored under the umbrella of traditional incentive/implementation/auction theory (see e.g. Yang and Huang (2005); Sandholm (2002); Wardrop (1952)). For example, as reported in Börjesson et al. (2012), experiments like the Stockholm congestion tax have shown that relatively small monetary incentives, in the form of small price variations over the duration of a day and the extent of the central business district, can successfully reduce congestion and gain popular acceptance.

There are also grounds for considering other influencing techniques, which rely on informational exchanges between the system and its users, rather than financial ones. These influencing techniques are increasingly important in transportation, especially as route guidance applications have become standard on most smartphone platforms. The applications act as route recommenders, and issue information on the monetary and travel time costs, rather than directly charging or paying users to influence behavior. These recommender systems are playing an increasing role in the daily life and decision–making process of commuters (Kearns et al. (2012)), and motivates our work.

In this paper, we investigate a model of informational nudging in a simplified context inspired by learning in repeated traffic games. While our model is not directly applicable to the transportation domain yet (because, in this case, each decision maker's reward – the experienced travel time – depends on the decisions of all others), it is appropriate to describe other decision maker/recommender interactions such as advertisement, propaganda, or ”true” recommendations about the quality of a good. Following the work of Saghai (2013); Thaler and Sunstein (2008), a *nudge* is defined as an influence which respects a decision–maker's choice set and is “substantially non–controlling,” in the sense that it is possible for the decision–maker to (*i*) become aware of the influence, (*ii*) inhibit its resulting propensity, and (*iii*) do so in an “effortless” manner. Thus, informational nudges have the desirable qualities of preserving the participant's freedom of choice, while making relatively minimal assumptions about her rationality, and requiring no money exchange between the system and its users.

Using the payoff–based learning model of Cominetti et al. (2010) (which represents a decision–maker as choosing her path according to a logit probability distribution dependent on announced payoff, and is thus consistent with some models of bounded rationality such as that of McKelvey and Palfrey (1995)), we introduce a recommender who provides possibly misleading payoffs for unchosen paths. As pointed out by Dworkin (2013), this kind of deception can be seen as a *bona fide* nudge provided conditions (*i*)–(*iii*) mentioned above can be

[⋆] This work was supported in part by an ISAE alumni grant and, in part, by the Office of the Vice–Chancellor for Research at UIUC, through the In3 Project “What Your Infrastructure Wants”. C.L. thanks California PATH at UC Berkeley for their hospitality while this work was initiated.

978-3-902823-55-7/2013 © IFAC

10.3182/20130925-2-DE-4044.00037

ascertained, and in particular, that a mechanism of reaction to deception is considered.

As a first step, we sidestep this element and study how each decision–maker's belief and behavior varies as a function of the nudge. We specifically consider the case when user payoffs are decoupled, which allows us to study the behavior of a single decision–maker.

After describing the model and its fluid limit in more detail in Section 2, we study simple nudging strategies in Section 3, in which the recommender uses constant lies to try and steer the system dynamics to a desired equilibrium. We show that it is possible for the recommender to choose nudges to create and locally stabilize any desired equilibrium in the fluid limit model, however, the equilibrium may not be globally asymptotically stable.

For the case when the user's choice set contains only two alternatives (Section 3.2), we fully characterize the set of equilibria of the fluid limit model that can be achieved via nudging. Moreover, we show that some equilibria cannot be created if the recommender is constrained to lie about only one alternative. This uncovers a maybe surprising framing effect, in the sense that the effect of lies are not symmetric: lying about a slow path being fast (in the traffic recommender example) has different effects than lying about a fast path being slow.

Finally, in Section 4, we discuss extensions and future work, including the presence of a recommender credibility feedback loop to address the substantial non–control conditions (*i*)-(*iii*).

2. MODEL DESCRIPTION

We consider a variation of the model originally proposed in Cominetti et al. (2010). A user repeatedly faces a set $\mathbf{A} = \{a_1, ..., a_K\}$ of K alternatives, each of which is associated with a true reward r_a.

After a choice is made at round n, an announced reward $w_a(n)$ is reported by the recommender for each alternative. These announced rewards depend on the user's choice and may differ from the true reward r_a due to the recommender's active influence. Based on the announced reward, the user updates her payoff perception vector x according to:

$$x_a(n+1) = (1-\gamma(n))x_a(n) + \gamma(n)w_a(n), \qquad (1)$$

for all a, where $\gamma(n)$ is a square summable, non-summable sequence, i.e., $\lim_{n\to+\infty}\gamma(n) = 0$, $\sum_{n=1}^{\infty}\gamma(n) = \infty$, and $\sum_{n=1}^{\infty}\gamma(n)^2 < \infty$. The user then chooses an alternative in the next round according to the probability distribution

$$Prob(\text{Alternative } a \text{ is chosen}) = \Pi_a(x) = \frac{e^{\beta x_a}}{\sum_{q=1}^{K} e^{\beta x_q}}. \qquad (2)$$

This is the well known logit choice model introduced by McFadden (1974) and used extensively in discrete choice theory. This model also makes similar assumptions about decision–making as parts of the literature on *bounded rationality* (in particular, the notion of quantal response equilibrium), and is thus consistent with our desire to capture a nudging–type influence. As explained by Cominetti et al. (2010), $\Pi(x) = (\Pi_1, ..., \Pi_K)$ can be thought of as maximizing $\sum_{a=1}^{a=K}(\Pi_a x_a - \frac{1}{\beta}\Pi_a(ln\Pi_a - 1))$, which is a regularization of the expected payoff by the entropy of the probability distribution. In turn, a large value of β corresponds to a user who mostly cares about maximizing expected payoff (in the limit of infinite β the probability distribution concentrates to pure actions with maximum x_a), while a small β indicates a relative desire for randomization. One should also note that logit rule implies $\Pi_a(x) \neq 1$ and $\Pi_a(x) \neq 0 \; \forall a \in \mathbf{A}$, which will be extensively used in subsequent derivations.

Yet another way to interpret model (1), (2), is to think of $\Pi_a(x)$ not as the probability that a *single* decision maker will choose alternative a, but as the fraction of a *population* of decision makers (all of which share the same intrinsic characteristics and desirability vector x for all alternatives) that will pick a.

This is consistent with McFadden's own justification of the logit choice model McFadden (1974), and allows one to also think of our proposed nudging strategy as a way for a recommender to modify the decisions of a group of decision makers, provided they all receive the same information from it.

We are interested in situations where the recommender may manipulate the announced rewards so as to eventually drive the user's choice to a desirable one, as defined by the recommender. Accordingly, we assume that the announced rewards take the form

$$w_a(n) = \begin{cases} r_a & \text{w.p. } \Pi_a(x(n)) \\ r_a + l_a(n) & \text{w.p. } (1-\Pi_a(x(n))). \end{cases} \qquad (3)$$

In words, this means that the recommender lies about the true reward for an alternative a by an amount $l_a(n)$ every time a is *not* chosen (and hence, presumably, every time it is impossible for the user to directly check the true value of the reward at the round n). Of course, such a scenario is not realistic if rewards $\{r_a\}$ are truly constant, since the user then just needs to choose every alternative once to uncover the value of the true reward, and has no reason to ever believe the recommender later on when a lie is announced. Note that it is also because these options are available to the user that the necessary conditions of "substantial non-controllability" are satisfied and that the mechanism considered here is a nudge. Thankfully, as explained e.g. by Borkar (2008), the analysis presented here is valid (and yields the same fluid limit and asymptotic behavior) if rewards are considered to be random variables rather than constant and deterministic.

This is explained in more detail after Proposition 1. From (1) and (3), we obtain the discrete stochastic difference equation

$$x_a(n+1) - x_a(n) = \begin{cases} \gamma(n)(r_a - x_a(n)) & \text{w.p. } \Pi_a\big(x(n)\big) \\ \gamma(n)(r_a + l_a(n) - x_a(n)) & \text{w.p. } 1-\Pi_a\big(x(n)\big) \end{cases} \qquad (4)$$

for all $a \in \mathbf{A}$.

The central question of interest, then, is whether there exists a strategy $\{l_a(n)\}_{a\in\mathbf{A}}^{n\geqslant 0}$ such that the dynamics described by (4) converge to a payoff vector x^* that corresponds to a desirable probability distribution vector π^*. In order to answer this question, we study the asymptotic behavior of (4) using the ODE / fluid method considered by Benaïm (1999), which takes advantage of the properties of $\{\gamma(n)\}$. More precisely, calling upon Corollary 5.4 by Benaïm (1999), we can state the following:

Proposition 1. Assume that the system of ordinary differential equations (ODE)

$$\begin{aligned}\frac{dx_a}{dt} &= \mathbb{E}(w_a(x) \mid \{x_a(s)\}_{s=0}^{n}) - x_a \\ &= r_a - x_a + l_a(x)(1-\Pi_a(x)) \forall a \in \mathbf{A} \qquad (5)\end{aligned}$$

admits a unique globally asymptotically stable equilibrium. Then the sequence $\{x(n)\}_{n\geq 0}$ of random vectors defined by (4) converges to this equilibrium almost surely.

We will henceforth mostly concern ourselves with ODE (5), with the understanding that stability results can only be interpreted directly in terms of $\{x(n)\}_{n\geq 0}$ in the case of global asymptotic stability. Before doing so, however, we note again that Proposition 1 still holds unchanged if, instead of representing mere deterministic and constant variables, each $r_a(n)$ and $l_a(n)$ is meant to designate the mean of a random variable of the form

$$\widetilde{r}_a(n) = r_a + \kappa_a(n) \qquad (6)$$

$$\widetilde{l}_a(n) = l_a(n) + \theta_a(n) \qquad (7)$$

where $\kappa_a(n)$ and $\theta_a(n)$ are independent martingales. In this case, the recommender's use of the random announced reward

$$\widetilde{w}_a(n) = \begin{cases} r_a + \kappa_a(n) & \text{w.p. } \Pi_a(x(n)) \\ r_a + l_a + \kappa_a(n) + \theta_a(n) & \text{w.p. } (1-\Pi_a(x(n))) \end{cases} \qquad (8)$$

in lieu of (3) yields to the same value of conditional expectation $\mathbb{E}(\widetilde{w}_a(n) \mid \{x_a(s)\}_{s=0}^{n})$ and, hence, same ODE as (5).

3. CONSTANT RECOMMENDER STRATEGIES

As explained above, the recommender's goal is to choose a strategy $l^* = (l_1^*, ..., l_K^*)$ such that dynamics (5) asymptotically converge to the set of payoff perception vectors corresponding to a desired probability distribution π^*. Note that this set is typically not a singleton, as can be seen, e.g., by considering the simple case of two alternatives, with $\pi^* = \left(\frac{1}{2}, \frac{1}{2}\right)$. In this case, every payoff perception vector of the form $\alpha(1,1)^T$, $\alpha \in \mathbb{R}$ produces the same distribution π^*. In this section, we focus on *constant* strategies that drive the dynamics to a single element of that set, although considering controls that make the full set a limit set of the dynamics would also be of interest. We consider such strategies first because they require very low attention on the recommender's part and, as we will show, can in some cases enforce almost any equilibrium effectively.

We first derive a condition on l^* that guarantees that an appropriate equilibrium is created, and then investigate additional requirements for its (global) stability.

3.1 Arbitrary number of alternatives

Proposition 2. System (5) admits an equilibrium corresponding to π^* under the constant strategy $l^* = (l_1^*, ... l_K^*)$ if and only if there exists a scalar $s \in \mathbb{R}$ and index i such that

$$l_a^*(n) = \frac{s - r_a + (1/\beta)ln(\pi_a^*/\pi_i^*)}{1-\pi_a^*} \qquad (9)$$

for all $a \in \mathbf{A}$ and all $n \geq 0$.

Proof. Assume that (9) holds for some s and i and define $\bar{x}$ by $\bar{x}_a = s + (1/\beta)ln(\pi_a^*/\pi_i^*)$ for all a. Then, we claim that $\bar{x}$ is an equilibrium. Indeed, notice that, for all a,

$$e^{\beta \bar{x}_a} = e^{\beta s}\frac{\pi_a^*}{\pi_i^*}. \qquad (10)$$

Hence, $\Pi_a(\bar{x}) = \frac{e^{\beta \bar{x}_a}}{\sum_j e^{\beta \bar{x}_j}} = \pi_a^*$, i.e., $\Pi(\bar{x}) = \pi^*$. From this, it is also clear that

$$r_a - \bar{x}_a + l_a^*(1-\Pi_a(\bar{x})) = 0$$

for all a, i.e., that $\bar{x}$ is an equilibrium. The converse implication can be shown from the same algebra, by essentially reversing the steps.

We now turn our attention to the local stability of a desired equilibrium. To this end, we compute the Jacobian $J(x)$ at that point, when a strategy of the form (9) is applied to the system. Noting that

$$\frac{d\Pi_a}{dx_q}(x) = \begin{cases} \beta\Pi_a(x)(1-\Pi_a(x)) & \text{if } q = a \\ -\beta\Pi_a(x)\Pi_q(x) & \text{if } q \neq a \end{cases}$$

find that

$$J(x) = \begin{bmatrix} -1-\beta l_1^*\pi_1^*(1-\pi_1^*) & ... & \beta l_1^*\pi_1^*\pi_K^* \\ \beta l_2^*\pi_2^*\pi_1^* & ... & \beta l_2^*\pi_2^*\pi_K^* \\ ... & ... & ... \\ ... & ... & ... \\ \beta l_K^*\pi_K^*\pi_1^* & ... & -1-\beta l_K^*\pi_K^*(1-\pi_K^*) \end{bmatrix} \qquad (11)$$

whenever equilibrium x corresponds to the desired probability distribution π^*. From this, we can use Gershgorin's theorem to derive the following sufficient conditions for local stability.

Proposition 3. If

$$\begin{cases} l_a^* & > -\dfrac{1}{\beta\pi_a^*(1-\pi_a^*)} \\ -\mid l_a^* \mid + l_a^* & > -\dfrac{1}{\beta\pi_a^*(1-\pi_a^*)} \end{cases} \qquad (12)$$

for all $a \in \{1, ..., A\}$, then the equilibrium corresponding to π^* is locally stable.

Proof. The first condition in (12) is equivalent to the center of every Gershgorin circle being located in the complex left half plane, while the second condition is equivalent to the radius of each circle being smaller than the distance between the center and the origin. Indeed,

$$\sum_{q=1, q\neq a}^{K} \mid J_{aq}(x) \mid = \beta \mid l_a^* \mid \pi_a^*(1-\pi_a^*) \qquad (13)$$

and thus the radius is smaller if and only if

$$|1+\beta^* l_a^*\pi_a^*(1-\pi_a^*)| > \beta|l_a^*|(1-\pi_a^*)\pi_a^*.$$

Taking into account the first condition in (12) then yields the stated inequality.

Combining Proposition 2 and 3, we can state the following achievability result.

Theorem 4. For every desired probability π^*, there exists a constant recommender strategy l^* (with $l_a^* > 0$ for all a) that creates and locally stabilizes an equilibrium corresponding to π^*.

Proof. Note that local stability condition (12) is always trivially satisfied if $l_a^* > 0$ for all a. It is thus enough to choose a strategy of the form (9) such that all lies are positive. This can be achieved by picking any $s > \max_a r_a$ and $i = \arg\min_j \pi_j^*$ in the characterization of Proposition 2.

With this characterization in hand, and in light of Proposition 1, it is natural to ask whether something more can be obtained, namely, whether global asymptotic stability of a desired equilibrium is achievable as well. Gershgorin's theorem can likewise be used to derive sufficient conditions toward answering this question.

Proposition 5. Let M be defined as $M = \max_{p \neq q} \mid l_p^* + l_q^* \mid$. Then, if

$$l_a^* > -\frac{4}{\beta} \text{ and } (-\frac{M}{2} + l_a^*) > -\frac{4}{\beta} \tag{14}$$

for all $a \in \mathbf{A}$, system (5) has a unique, globally asymptotically stable equilibrium corresponding to $\pi^* = (\pi_1^*, ..., \pi_K^*)$, to which the sequence $\{x(n)\}_{n \geq 0}$ defined by (4) also converges almost surely.

Proof. From contraction theory (Jouffroy and Slotine (2004)), we know that a sufficient condition for the equilibrium point of (5) to be globally asymptotically stable is the existence of $\delta > 0$ such that the symmetric part of the Jacobian satisfy

$$J_{sym}(x) = \frac{1}{2}(J^T(x) + J(x)) \preceq -\delta I \text{ for all } x.$$

Using Gershgorin's theorem again, a sufficient condition for this to hold is

$$l_a^* > -\frac{1}{\beta \Pi_a(x)(1 - \Pi_a(x))}$$

and

$$1 + \beta l_a^* \Pi_a(x)(1 - \Pi_a(x)) > \frac{M\beta \Pi_a(x)(1 - \Pi_a(x))}{2}$$

for all a and x, where we have used the fact that

$$\sum_{q=1, q \neq a}^{K} \mid [J_{sym}(x)]_{aq} \mid < \frac{M\beta \Pi_a(x)(1 - \Pi_a(x))}{2}, \ \forall a$$

in the second inequality. Now, noting that $\Pi_a(x)(1 - \Pi_a(x)) < \frac{1}{4}$ for all x and a, we see that (14) is a sufficient condition for this latter inequality to hold.

3.2 The two alternative case

The previous section presented sufficient conditions for stability in the general case, and showed that every desirable equilibrium can be created and locally stabilized with a constant recommender strategy. In this section we focus on the simpler case of two alternatives ($K = 2$), and characterize the full set of asymptotic behaviors achievable with constant recommender strategies. In particular, we show that a poor choice of lies can result in unstable equilibria, which has implications for robustness of these strategies.

When only two choices are offered to the user the differential system reduces to

$$\begin{bmatrix} \dot{x_1} \\ \dot{x_2} \end{bmatrix} = \begin{bmatrix} -x_1 + r_1 + l_1 \left(1 - \frac{1}{1 + e^{\beta(x_2 - x_1)}}\right) \\ -x_2 + r_2 + l_2 \frac{1}{1 + e^{\beta(x_2 - x_1)}}. \end{bmatrix}$$

The study of this system can be simplified further by considering the differences $x_2 - x_1$. Defining $X_{21} = x_2 - x_1$ and $R_{21} = r_2 - r_1$, the dynamic of X_{21} can be reduced to

$$\frac{dX_{21}}{dt} = R_{21} - X_{21} - l_1 + \frac{l_1 + l_2}{1 + e^{\beta X_{21}}} \tag{15}$$

Case $l_1 = 0$ As a first step and in order to reveal an assymetry between the two alternatives we consider the situation where the recommender is restricted to lie only about the second alternative, i.e., $l_1 = 0$. In $l_1 = 0$ case, (15) can be further simplified to

$$\frac{dX_{21}}{dt} = R_{21} - X_{21} + \frac{l_2}{1 + e^{\beta X_{21}}} \tag{16}$$

In this case, the number and nature of equilibria is determined by the solutions of $f(X, l_2) = -R_{21}$, where function f is defined by

$$f(X, l_2) = \frac{l_2}{1 + e^{\beta X}} - X$$

for all X. A rapid study of the function reveals the following:

Lemma 6. (*i*) For every value of l_2, $\lim_{X \to +\infty} f(X, l_2) = -\infty$ and $\lim_{X \to -\infty} f(X, l_2) = +\infty$.

(*ii*) When $l_2 \geq -\frac{4}{\beta}$, function $f(\cdot, l_2)$ is monotonically decreasing.

(*iii*) When $l_2 < -\frac{4}{\beta}$, function $f(\cdot, l_2)$ admits a local minimum X_1^* and a local maximum X_2^*.

Proof. Let $g_{l_2}(\cdot) = f(\cdot, l_2)$ and define the variable $Y = e^{\beta X}$ so that

$$g'_{l_2}(X) = -\frac{Y^2 + (2 + \beta l_2)Y + 1}{(1 + Y)^2}.$$

The sign of the derivative is given by the opposite of the sign of the polynomial $P(Y) = Y^2 + (2 + \beta l_2)Y + 1$ whose discriminant is $\beta l_2(4 + \beta l_2)$. Hence, when $-\frac{4}{\beta} \leq l_2 \leq 0$, P does not vanish and is positive for all Y. When $l_2 < -\frac{4}{\beta}$ or $l_2 > 0$, P has two roots

$$Y_1^* = (1/2)\left(-2 - \beta l_2 - \sqrt{\beta l_2(4 + \beta l_2)}\right)$$
$$Y_2^* = (1/2)\left(-2 - \beta l_2 + \sqrt{\beta l_2(4 + \beta l_2)}\right)$$

and is negative on the interval $[Y_1^*, Y_2^*]$. However, observe that these roots are either both positive, when $l_2 < -\frac{2}{\beta}$ or both negative, when $l_2 \geq -\frac{2}{\beta}$. Hence, $g'_{l_2}(X)$ only changes sign when $l_2 < -\frac{4}{\beta}$, in which case $f(\cdot, l_2)$ admits a local minimum at $X_1^* = \frac{1}{\beta} ln Y_1^*$ and local maximum at $X_2^* = \frac{1}{\beta} ln Y_2^*$.

From item (*ii*) in Lemma 6 and the discussion preceding it, it follows that system (15) admits a unique globally asymptotically stable equilibrium whenever $l_2 > -\frac{4}{\beta}$, independently of the value of R_{21}. For this reason, this will be called the *unconditional stability* region. This is in agreement with sufficient condition (14) applied with $K = 2$, $l_1 = 0$. However, we also see that global asymptotic stability can be achieved for other values of l_2, depending on the value of R_{21}. Indeed, from the discussion above, we have the following

Theorem 7. Let Ω be the subset of the (l_2, R_{21})–plane defined by

$$\Omega = \{(l_2, R_{21}) \mid l_2 < -\frac{4}{\beta}, g_{l_2}(X_2^*) > -R_{21} > g_{l_2}(X_1^*)\},$$

where

$$g_{l_2}(X_1^*) = \frac{2l_2}{\beta l_2 + \sqrt{\beta l_2}(4+\beta l_2)} + \frac{ln\left[-1-\beta l_2/2-(1/2)\sqrt{\beta l_2(4+\beta l_2)}\right]}{\beta}$$

$$g_{l_2}(X_2^*) = -(\beta l_2 + \sqrt{\beta l_2(4+\beta l_2)} + ln(4) - 2ln\left(-2-\beta l_2+\sqrt{\beta l_2(4+\beta l_2)}\right))/(2\beta).$$

System (15) admits:

- three equilibrium points (two stable and one unstable one), if and only if $(l_2, R_{21}) \in \Omega$,
- two equilibrium points (one stable and one saddle–node), if and only if $(l_2, R_{21}) \in \partial\Omega$,
- a single globally asymptotically stable equilibrium otherwise.

A picture of set Ω is provided in Figure 1. It can be shown that the lower part of the boundary is the graph of a concave decreasing function of l_2, while the upper part is the graph of a convex decreasing function of l_2.

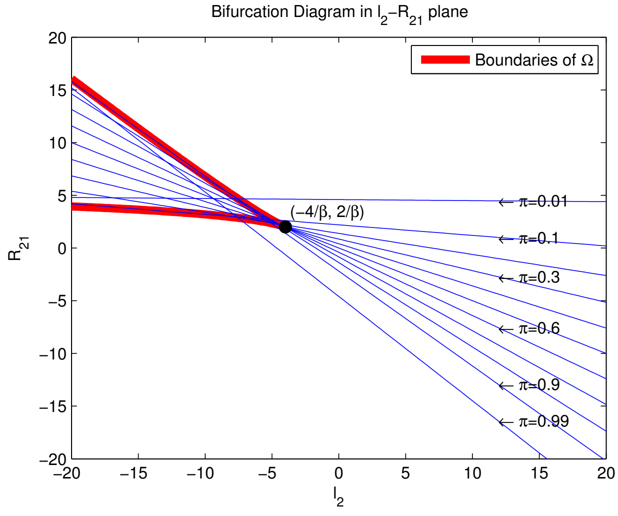

Fig. 1. Iso π and Ω region

Going back to the characterization (9), one can see that the only constant strategy inducing equilibrium $(\pi_1^*, 1-\pi_1^*)$ with $l_1^* = 0$ is such that $l_2^* = -\frac{R_{21}}{\pi_1^*} + \frac{1}{\beta\pi_1^*} ln(\frac{1-\pi_1^*}{\pi_1^*})$. This means that all the points located on the same straight (blue) line in Figure 1 yield the same equilibrium probability distribution. For this reason, we refer to each of these curves as an iso–(equilibrium) probability or iso–π line.

In addition, note that for a given value of $\pi_1^* \neq 0$ and R_{21}, there is a unique choice of l_2^* imposing the probability distribution $(\pi_1^*, 1-\pi_1^*)$ in equilibrium, namely, the abscissa of the unique intersection point between the corresponding iso–π and "$y = R_{21}$" line in the (l_2, R_{21})–plane. As is apparent in Figure 1, this intersection lies in region Ω for some values of π_1^* and R_{21} (e.g., $\pi_1^* = 0.3$ and $R_{21} = 10$) and, hence it is impossible to globally asymptotically stabilize the corresponding probability distribution with a constant strategy.

From this discussion, it follows that constant recommender strategies with $l_1^* = 0$ are not always satisfactory, because there exist values of R_{21} for which a range of desired probability distributions cannot be globally stabilized by them. Note, however, that all such values of $R_{21} = r_2 - r_1$ are positive. This means that it may not be possible for the recommender (using constant strategies) to drive the user to a state where she prefers the least favorable alternative by misrepresenting only the value of most favorable alternative. However, it is possible to attain a state where the least favorable alternative is preferred by lying only about this alternative, regardless of the values of r_2 and r_1. In other words, (and to use an analogy more appropriate for Yelp than, Google Navigation, our original motivation) *"lying about 'Gourmet Restaurant' being 'Fast Food' may not work, but lying about 'Fast Food' being 'Gourmet Restaurant' does"!* Our model shows that in the situation where $r_2 \gg r_1$ enforcing the desired equilibrium is impossible. When recommender computes the nudge l_2 he wants to create π_1^* (corresponding to X_{21}^*) the equilibrium he thinks is desirable. But when analyzing the ODE it turns out that the equilibrium point π_1^* is unstable, and that the recommender creates two other stable equilibria $\pi_1^{*-} < \pi_1^*$, $\pi_1^{*+} > \pi_1^*$ (corresponding to $X_{21}^{*-} > X_{21}^*$ and $X_{21}^{*+} < X_{21}^*$. Hence the equilibrium points that are reached depends on the initial value of the user payoff estimate X_{21}^0

(1) If $X_{21}^0 > X_{21}^*$ the user reaches π_1^{*-}
(2) If $X_{21}^0 < X_{21}^*$ the user reaches π_1^{*+}

Hence when the true payoff difference is high enough the recommender polarizes the opinions. Some of the users are not going to be influenced enough to change their choice while others will be confortced by the nusges.

In the next section, we show that allowing the recommender to lie about both alternatives removes this difficulty.

General case Let us now consider constant strategies in which the value of l_1 is free. Note that system (15) with $l_1 \neq 0$ can be put in the form of (16) when replacing R_{21} and l_2 in this latter equation by

$$\widetilde{R}_{21} = R_{21} - l_1 \quad (17)$$
$$\widetilde{l_2} = l_1 + l_2. \quad (18)$$

In other words, the dynamics induced by this strategy on a problem with given value R_{21} are the same as those induced by a strategy with no lie permitted on the fist alternative and $\tilde{l}_2$, on a system with given value $\tilde{R}_{21}$.

As a result, the same analysis as in Section 3.2.1 can be carried out in the $(\tilde{l}, \tilde{R}_{21})$–plane. From this, we see that if, for a given R_{21}, l_1 is chosen so that

$$\tilde{l}_2 = l_1 + l_2 \geq -\frac{4}{\beta}$$

and

$$\tilde{R}_{21} = R_{21} - l_1 = -\pi_1^* \tilde{l}_2 + \frac{1}{\beta} ln(\frac{1-\pi_1^*}{\pi_1^*}),$$

for example, then the equilibrium $(\pi_1^*, 1-\pi_1^*)$ is created and globally stabilized. Clearly, these conditions can always be satisfied for fixed R_{21} and π^*, since both l_1 are l_2 are free parameters. In fact, one can even find the smallest lie in the unconditional stability region by solving the following quadratic program:

$$\min \quad l_1^2 + l_2^2 \tag{19}$$

$$s.t. \quad l_1 + l_2 \geq -\frac{4}{\beta} \tag{20}$$

$$R_{21} = l_1 - \pi_1^*(l_2 + l_1) + \frac{1}{\beta} ln \left(\frac{1 - \pi_1^*}{\pi_1^*} \right), \tag{21}$$

which is always feasible.

4. CONCLUSION, DISCUSSION, AND FUTURE WORKS

In this paper, we considered a variation of the payoff–based learning model of Cominetti et al. (2010) in which the recommender can actively and strategically modify announced rewards for unchosen alternatives, so as to eventually induce the user to make particular choices. We considered constant lying strategies and showed that they can be effective in globally stabilizing an equilibrium corresponding to any desirable probability distribution. In so doing, we also showed that there is an asymmetry between the effects of lies.

Several directions still need to be explored. First, staying within the confines of the model presented here, one might want to consider more general recommender strategies than the constant ones, maybe as a way to achieve better convergence rates in closed–loop. Indeed, it is straightforward to see that a feedback strategy of the form

$$l_a^*(x) = \frac{x_a^* - r_a}{(1 - \Pi_a(x))}, \tag{22}$$

where x^* is chosen such that $\Pi_a(x^*) = \pi_a^*$ for all a globally asymptotically stabilizes the equilibrium x^*, since the closed–loop system then is

$$\frac{dx_a}{dt} = x_a^* - x_a \quad \text{for all } a.$$

However, this feedback strategy requires the recommender to access either the user's perception vector x or the probabilities $\{\Pi_a(x)\}$, both of which are unlikely to be available in practice (a surrogate for $\Pi_a(x)$ might be obtained by monitoring the empirical distribution of the user's choices up to decision time as is done, e.g., in fictitious play. However, this introduces mistakes in the control strategy and requires us to further study its robustness). In addition, strategy (22) has the drawback of requiring very large lies for initially rarely chosen alternatives, even if the equilibrium value of their corresponding probability distribution is large – a property not shared by constant strategies, and inconsistent with the conditions underlying the substantial non–controlling aspect of a nudge. This is akin to the issue of large gains in classical control theory and certainly deserves further investigation.

Another avenue of current work is the incorporation of a notion of *credibility* to the present model, whereby the user reacts to false announced rewards by updating an additional trust vector in a manner similar to (4). As explained before, this is central to being able to consider the control strategies proposed here as *bona fide* nudges.

Finally, in connection with our original traffic route choice motivation, we are also considering extensions of the present idea of recommender lies as control strategies, and of the present results, to situations involving multiple users and where rewards to a user depend on her previous actions, as well as on those of other users.

REFERENCES

Benaïm, M. (1999). Dynamics of stochastic approximation algorithms. *Seminaire de probabilites XXXIII*, 1–68.

Börjesson, M., Eliasson, J., Hugosson, M.B., and Brundell-Freij, K. (2012). The Stockholm congestion charges–5 years on. Effects, acceptability and lessons learnt. *Transport Policy*, 20, 1–12.

Borkar, V.S. (2008). *Stochastic approximation: a dynamical systems viewpoint.* Cambridge University Press, Cambridge.

Cominetti, R., Melo, E., and Sorin, S. (2010). A payoff-based learning procedure and its application to traffic games. *Games and Economic Behavior*, 70(1), 71–83.

Dworkin, G. (2013). Lying and nudging. *Journal of Medical Ethics.*

Jouffroy, J. and Slotine, J.J. (2004). Methodological remarks on contraction theory. In *Proceedings of the 43rd IEEE Conference on Decision and Control*, 2537–2543.

Kearns, M., Pai, M.M., Roth, A., and Ullman, J. (2012). Private equilibrium release, large games, and no-regret learning. Unpublished manuscript arXiv:1207.4084[cs.GT].

McFadden, D. (1974). *Frontiers in Econometrics.* Academic Press, New York.

McKelvey, R.D. and Palfrey, T.R. (1995). Quantal response equilibria for normal form games. *Games and economic behavior*, 10(1), 6–38.

Saghai, Y. (2013). Salvaging the concept of nudge. *Journal of Medical Ethics.*

Sandholm, W.H. (2002). Evolutionary implementation and congestion pricing. *Review of Economic Studies*, 69(3), 667–689.

Thaler, R.H. and Sunstein, C.R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* Yale University Press.

Wardrop, J.G. (1952). Some theoretical aspects of road traffic research. In *Road Engineering Division Meeting*, volume 1, 325–362.

Yang, H. and Huang, H.J. (2005). *Mathematical and economic theory of road pricing.* Elsevier, Oxford.