



A GAME OF ASYMMETRIC INFORMATION BETWEEN A THERAPIST AND A MENTALLY ILL PATIENT

KJETIL K. HAUGEN AND SOLFRID VATNE

Abstract. In this paper, we follow up research of Vatne [23] related to how mentally ill patients could (and perhaps should) be treated. We propose, model, and analyze a simple sequential game of incomplete and asymmetric information where the patient moves first, signaling behavior which is observed by the therapist with lack or limited knowledge of the actual patient type. We assume greatly approximated strategy spaces both for patient and therapist, but will still claim that our results increase knowledge of this important relation. Knowledge that may be critically important in continued improved treatment of patients with mental illness.

1. INTRODUCTION, MOTIVATION AND OUTLINE

1.1. Introduction

It is a well-known fact that both diagnosis and efficient treatment of mentally ill patients is difficult. So far, medical science has struggled both with categorizing diseases as well as with providing treatment which with relative certainty will cure or ease patients' symptoms. So far, even though most experts seem to agree on the opposite, much treatment is performed through medication – often by drugs with heavy adverse effects.

If such a description (even though obviously biased) is somewhat close to reality, one would expect a problem of the following type. If the therapist is uncertain on what type of patient he (or she) is treating, the potential for successful treatment is limited. In many instances, one should hence expect that wrong treatment is prescribed for many patients with strong negative outcomes for the individual patient and, in a longer perspective, for society. International literature, focusing on patients' views illustrates such a perspective [6, 12]. Additionally, the fact that the therapist (and the system) may have incentives and power towards calm patients (a typical effect of heavy medication) makes the situation even worse.

Vatne's research [20–23] is concentrated on such topics. In her studies of milieu-therapists' limit-setting in mental health, she uses the terms *Acknowledging* as opposed to *Corrective* actions – symbolizing a simplification of the “treatment space”. She argues that extensive use of corrective treatment may be the biggest obstacle in obtaining a constructive cooperation between patient and therapist, and that the corrective action itself constructs conflicts which again may lead to

MSC (2010): primary 91A28, 91A80; secondary 62-00.

Keywords: game theory, incomplete and asymmetric information, mental illness.

more necessary (or better unnecessary) corrective action - a viscous circle. The following two observed narratives from Vatne's research illustrate such a situation:

1) The narrative of correcting uncleanness

Ann, the patient in this narrative, a 60-year-old woman diagnosed with Schizophrenia, was reported to have declined to clean herself in 14 days. The responsible nurse decides for herself that this day Ann has to take a shower because she smells and can get sore. When we enter Ann's room, the nurse asks Ann if she wants to take a shower before or after breakfast. Ann says she has already cleaned herself and that it now is time for breakfast. After the meal, the nurse tries to motivate Ann to shower, by explaining how important it is to take care of her good skin and well-being. She walks over to Ann's wardrobe to get her some new clean clothes. Ann screams: "Don't open my closet!" The nurse's answer is that she feels pity for how messy Ann looks. Ann then points up to the ceiling, saying: "It is dirty here too, somebody should come and clean here as well." "You should think more about your own appearance, you look negligent", the nurse answers. The next half hour, the nurse tries to make an agreement with Ann taking a shower by seducing her in varied ways: offers coffee, cakes and smoke, and warns her that the cakes might quickly be eaten by other patients. Ann mumbles: "I am out of control." After a while Ann walks to the bathroom. The nurse following Ann says: "When you now are nearly undressed, I will wash you below, and put on some lotion." Ann bursts out: "You are making me even crazier!" When we meet Ann the next day, Ann tells the nurse: "You are a bad nurse. You don't deserve any cake and coffee either!" Afterwards, the nurse's reflection was that Ann's reaction was not expected and that Ann probably was too intelligent and saw through her way of limit-setting.

In the literature, such a correcting process; containing a continuum of interventions from informing, motivating, reality orienting and fictitious agreements to coercion, are described to be therapeutic [7,9]. However, the effect on the patient is being out of control and humiliating, which is counteracting experiences of therapy.

2) A patient's unforeseen aggressive outburst on "normal caring"

Torill has a serious mental illness and is hospitalized in an acute psychiatric ward due to aggravated symptoms. She has been confined to her room against her will as a result of her current mental condition and sometimes aggressive behavior. Various milieu-therapists are responsible for Torill while she is isolated in her room. Torill has several times tried to escape from her room to be with the other patients in the living room, especially during the meals. She says she likes to eat together with other people. When today's afternoon coffee with cakes is served, they forget Torill. Torill asks for the coffee, and her responsible nurse leaves

the isolate to fetch it. Since there are no cakes left, she returns with a sandwich. Torill responds with anger and frustration and accuses the nurse of denying her even a measly piece of cake. Torill then runs out of the room to look for the cakes herself. When she returns without cakes, the nurse tries to explain what had happened. But Torill is still angry and confronts the nurse with a torrent of accusations. She speaks and acts in a threatening manner, and, with her mouth full of food, spits at the nurse. That results in the nurse withdrawing and leaving the room. The nurse's reflection was that the patient's reaction on a sandwich as a substitute to a cake was totally unexpected and she was not able to cope with the situation.

Vatne discusses and explains the concept at hand in the latter narrative – a weak correcting (or acknowledging) intervention in [23]. An acknowledging intervention to information could be to say: I can understand if it is so that you felt offended while you were offered a sandwich instead of a cake...?

1.2. Motivation

The main motivation behind this study is an idea of approaching this problem from a (presumably) novel and different perspective – game theory. Game theory offers a set of tools to investigate, analyze and sometimes predict outcomes of situations involving decision making among more than one agent, where the utility of agents is constructed so that individual utility is a function of other agents' as well as own decisions. More specifically, games of incomplete information, offer tools which cope with situations where agents' individual pay-offs (e.g., types) may be unknown to other agents, a situation clearly resembling the situation briefly discussed in Subsection 1.1.

1.3. Outline

In Section 2, we investigate and discuss relevant literature with emphasis on applications of game theory related to treatment of mentally ill patients. In Section 3 we propose our game framework and discuss its benefits and limitations. The model is analyzed in Section 4, while concluding remarks and suggestions for further research are left for Section 5.

2. RELEVANT LITERATURE

Game Theory (GT) has been (although not primarily) a focus of interest in the field of health research since 1960. The theory has especially been applied to themes such as risk analysis in health care costs studies [1, 14, 16] and management research [5, 10]. Topics more close to health phenomenon as for instance medical ethics are treated in [15] while various psychological issues as rationality in social interaction, cooperation and medical consultation can be observed in [3, 4, 18]. However, all those articles are mainly using GT from a theoretical perspective. Research has mainly been focusing on how GT contributes in explaining empirically observed phenomena, e.g., cooperation in antisocial personality disorder [4]

and primary care [17], and is not intended to provide a test of predictions based on GT. Some recent interesting microbiological studies present GT as a basis for modeling systems of interaction between tumor cells [2, 11, 19], but do not seem very relevant in our setting. As a consequence, to the best of our knowledge, we have found no systematic accounts of Game Theory used to analyze a sequential game of incomplete information in a patient-provider relationship.

3. MODELING

We adopt a modeling structure much in the original spirit of Harsanyi [8], perhaps more modernly described in Rasmusen's *Follow the Leader III* [13].

We define two players: the therapist (\mathcal{T}) and the patient (\mathcal{P}). The players engage in a two-stage sequential game where \mathcal{P} moves first. That is, we assume that the patient makes a signal of say Cooperative (Passive - P) nature or alternatively more Opposing (Aggressive - A). Hence, we define the set of actions for player \mathcal{P} as $\{A, P\}$. Likewise (for simplicity), we restrict the actions for the therapist to be Corrective (C) or Recognizing/Acknowledging (R)¹. So, the action set for player \mathcal{T} are chosen in $\{C, R\}$.

Furthermore, we assume that \mathcal{T} has prior information of the patient type defined by a three-point probability density; say $\mathbf{p} = [p_1, p_2, 1-p_1-p_2]$ where $p_1, p_2, 1-p_1-p_2$ are probabilities of the patient being of types 1, 2 or 3 respectively. It seems reasonable to interpret these three patient-types as different degrees of illness, where type 1 could be mild, 2 is average while 3 is significantly more severe.

Introducing general pay-offs for the two players (a_j^i for \mathcal{T} and b_j^i for \mathcal{P} where $j \in \{1, 2, 3, 4, 5, 6\}$ and $i \in \{1, 2, 3\}$) the gaming situation (without incomplete information) may be illustrated as shown in Figure 1.

In Figure 2, the game is shown after the *Harsanyi Transformation*.

In Figure 2, the game is transformed from a game of incomplete information into a game of complete but imperfect information. Nature \mathcal{N} is introduced as a player, making his move (choosing the patient type 1, 2 or 3) unobserved by the therapist. This can be observed by the dashed information sets indicating that the therapist does not know in which of the nodes $\mathcal{T}_1, \mathcal{T}_3, \mathcal{T}_5$ or $\mathcal{T}_2, \mathcal{T}_4, \mathcal{T}_6$ he is. Note also that the patient knows who he is, hence this is a game of asymmetric information where the patient knows his own type, while the therapist is uninformed on the patient type. One could of course question such an assumption, certain situations might obviously call for other models including less rational patients and better informed therapists. However, we stick to this formulation in subsequent paragraphs.

4. ANALYSIS

Analysing incomplete information games can be done in various ways. The classical (Harsanyi) way includes "guessing" on a Nash Equilibrium (NE) and computing (Bayes) best replies in order to confirm NE existence. In this case, this method seems appropriate. The reason is straightforward. Underlying the problem is an

¹Obviously, both the patient and the therapist will have much wider action spaces to operate on in the real world. However, we deliberately try to simplify as much as possible in order to get a traceable game to analyze.

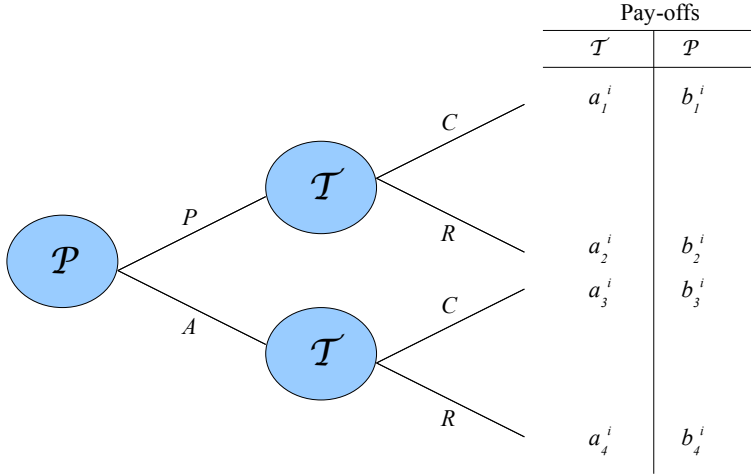


Figure 1. The game *before* the Harsanyi Transformation.

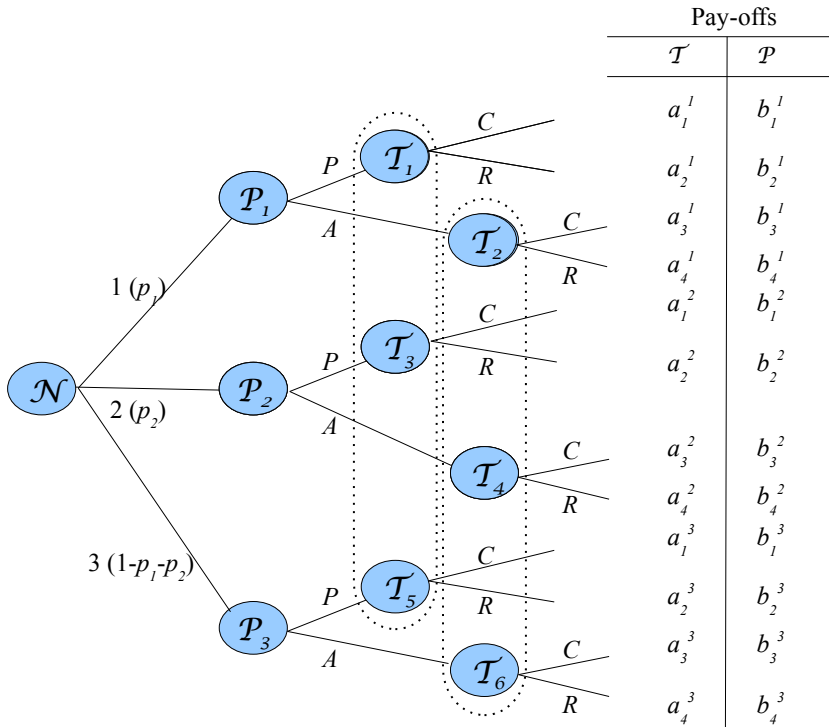


Figure 2. The game *after* the Harsanyi Transformation.

implicit hypothesis related to misuse of therapist corrective action. The main point in Vatne's research [20–23], is focused on such a hypothesis. In addition, Vatne points out real world cases which strongly support such a hypothesis. As a consequence, we are mainly interested in what forces might drive such NEs. That is, what values of model inputs (p and a_j^i, b_j^i) drives strongly corrective NEs. For instance; given a pay-off structure (given values on a_j^i, b_j^i), what patient types receive corrective treatment and vice versa. Alternatively; given a certain patient type, what therapist regimes tend to choose corrective treatment. Hopefully, answering such questions may shed light on how treatment should be organized in order to minimize corrective treatment. Another line of thinking could be to define various groups of a_j^i, b_j^i s, symbolizing different existing therapeutic regimes and check how these regimes react to different patient types by computing NEs as a function of patient type p .

Ultimately, we aim to reach conclusions related to how the pay-off structure must be defined in order to minimize (or reach more acceptable) levels of corrective action. This, as opposed to a classical prohibitive system. That is, instead of defining what is allowed and what not, we aim to establish a mechanism (how must a_j^i s and b_j^i s relate to each other) in order to achieve certain targets related to use (misuse) of corrective treatment action.

Anyway, we start out trying to keep the input set p, a_j^i, b_j^i as general as possible to keep the analytic space as open as possible.

Before the actual analysis, it is meaningful to outline all possible (pure strategy) NEs for the game defined in Figures 1 and 2. The complete set (through enumerating possible combinations) is shown in Tables 1 and 2. In Table 1,

NE	Label
$(C A, R P)$	DEFAULT
$(C A, C P)$	OBSERVED
$(R A, R P)$	UTOPIA
$(R A, C P)$	IMPOSSIBLE

Table 1. Possible pure strategy Nash Equilibrium strategies for the therapist.

NE	Label
$(P 1, P 2, P 3)$	EASY
$(A 1, A 2, A 3)$	DIFFICULT
$(P 1, P 2, A 3)$	EXPECTED I
$(P 1, A 2, A 3)$	EXPECTED II
$(P 1, A 2, P 3)$	
$(A 1, P 2, P 3)$	
$(A 1, P 2, A 3)$	
$(A 1, A 2, P 3)$	

Table 2. Possible pure strategy Nash Equilibrium strategies for the patient.

the notation $X|Y$ means that the therapist responds with X after observing the patient's action Y ². Furthermore, we have labeled each of these four pure strategy NEs with labels; DEFAULT, OBSERVED, UTOPIA and IMPOSSIBLE.

The first label, DEFAULT, corresponds to the situation as we would expect the system to be designed for. That is, difficult patients are treated with corrective action, while less difficult patients are treated with acknowledging action. All other three labels can to some extent be viewed as anomalies, The label OBSERVED refers to a situation we “are afraid” we observe to often, UTOPIA, where any patient type is treated nicely (some might state that this is the objective) and finally, a reversed default or IMPOSSIBLE NE where aggressive patients are treated with ease while cooperative patients are treated harshly.

Table 2 contains similar information for the patient. Now, we can (in principle) combine any element in Tables 1 and 2 to produce 32 possible pure strategy NEs. Half of the NE strategies in Table 2 are of course silly. The last four ones, under the mid horizontal line, are in conflict with our initial definition of patient behavior. A patient acting more aggressively in an illness state, defined as less severe makes no sense. Additionally, the two first lines of Table 2 are plausible, but not very interesting in this setting. After all, we are aiming to study the conflict related to therapeutic actions in the action set acknowledging or correcting, and a patient either responding only aggressively or passively no matter who he is seems uninteresting in this setting. The candidates labeled EXPECTED I and EXPECTED II are both relevant candidates for analysis. However, to keep the complexity at a reasonable level we stick to a single expected patient behavior unless anything else is explicitly stated. That is, when we in subsequent paragraphs mention the (OBSERVED) NE, we actually mean the (EXPECTED I; OBSERVED) NE.

The aim of the forthcoming analysis, is hence to investigate which parametric sets lead to which of the four defined NE-categories in Table 1.

We start by analyzing the (EXPECTED I; DEFAULT) NE. This construct may be interpreted as a situation where patients are treated and react as they “should”. It is appropriate to write out the actual meaning of this NE

$$\underbrace{(P|1, P|2, A|3)}_{\mathcal{P}}; \overbrace{C|A, R|P}^{\mathcal{T}}. \tag{4.1}$$

As (4.1) indicates, the tractable patient (\mathcal{P}) reacts passively (P) if he is of type 1 or 2, while he reacts aggressively (A) if he is of type 3. The therapist (\mathcal{T}) treats the patient correctively (C) after observing A while acknowledging (or recognizing) treatment is implemented if the therapist observes (P).

The computations involved in securing (4.1) as an NE starts by calculating so called Bayesian belief updates. Bayes law in our notation can be expressed as

$$Prob(1|P) = \frac{Prob(P|1) \cdot Prob(1)}{\sum_{j=1}^3 Prob(P|j) \cdot Prob(j)}. \tag{4.2}$$

²In any incomplete information game, we must look for so called contingent equilibrium strategies. That is, the therapist should treat the patient differently depending on patient type, which of course is reasonable. This does not rule out the possibility of treating the patient independently of patient type as NEs OBSERVED and UTOPIA indicate.

Given our proposed NE (4.1), it is straightforwardly deduced that $Prob(P|1) = Prob(P|2) = 1$ and $Prob(P|3) = 0$. Then, it follows from (4.2) that

$$Prob(1|P) = \frac{1 \cdot p_1}{1 \cdot p_1 + 1 \cdot p_2 + 0 \cdot (1 - (p_1 + p_2))} = \frac{p_1}{p_1 + p_2}.$$

Now, as $Prob(P|3) = 0$, the same holds for $Prob(3|P)$ and $Prob(2|P) = 1 - Prob(1|P)$ or

$$Prob(2|P) = 1 - \frac{p_1}{p_1 + p_2} = \frac{p_2}{p_1 + p_2}.$$

Then, we move to examining best replies. If the therapist observes P , he can choose actions C or R .

1) Choosing C

Expected pay-off for the therapist choosing C after observing P , $E_{\mathcal{T}}(C|P)$, is

$$E_{\mathcal{T}}(C|P) = a_1^1 \cdot \frac{p_1}{p_1 + p_2} + a_1^2 \cdot \frac{p_2}{p_1 + p_2}.$$

2) Choosing R

Expected pay-off for the therapist choosing R after observing P , $E_{\mathcal{T}}(R|P)$, is

$$E_{\mathcal{T}}(R|P) = a_2^1 \cdot \frac{p_1}{p_1 + p_2} + a_2^2 \cdot \frac{p_2}{p_1 + p_2}.$$

Now, to secure our prescribed NE (4.1), the therapist must choose R after observing P ($P|1, P|2, A|3; C|A, \mathbf{R}|\mathbf{P}$). Hence, $E_{\mathcal{T}}(R|P) > E_{\mathcal{T}}(C|P)$ or

$$a_2^1 \cdot \frac{p_1}{p_1 + p_2} + a_2^2 \cdot \frac{p_2}{p_1 + p_2} > a_1^1 \cdot \frac{p_1}{p_1 + p_2} + a_1^2 \cdot \frac{p_2}{p_1 + p_2}. \quad (4.3)$$

As $p_1 + p_2 > 0$, we can rewrite (4.3) as

$$a_2^1 \cdot p_1 + a_2^2 \cdot p_2 > a_1^1 \cdot p_1 + a_1^2 \cdot p_2. \quad (4.4)$$

Consequently, the constraint (4.4) must be satisfied to secure the prescribed NE (4.1).

However, this is not the only constraint needed. Let us find the remaining ones. The next step is to analyze \mathcal{T} 's best response to observing A . This situation is easier (deterministic) as $Prob(1|A) = Prob(2|A) = 0$ and $Prob(3|A) = 1$. The point here, is to secure that the $C|A$ -outcome is realized. Given that \mathcal{T} believes the state of nature is 3, his best response to A should be C . This is easily secured (directly from the Harsanyi-transformed game of Figure 2) by

$$a_3^3 > a_4^3. \quad (4.5)$$

At this point, we have shown that \mathcal{T} 's strategy ($C|A, R|P$) is his best response to \mathcal{P} 's strategy ($P|1, P|2, A|3$). Now in order to fully secure the prescribed NE (4.1), we also need to demonstrate (or to be more precise, find conditions for) that \mathcal{P} 's strategy ($P|1, P|2, A|3$) is his best response to \mathcal{T} 's strategy ($C|A, R|P$).

Note (from Figure 2) that \mathcal{P} can choose either P or A in any of the nodes \mathcal{P}_1 , \mathcal{P}_2 or \mathcal{P}_3 . It is straightforward to run through and find best replies for \mathcal{P} , given the prescribed NE, as follows:

In $\mathcal{P}_1, \mathcal{P}$ chooses between

$$\begin{cases} P, \text{ knowing that } \mathcal{T}_1 \text{ chooses } R, \text{ pay-off must be } b_2^1, \\ \text{or} \\ A, \text{ knowing that } \mathcal{T}_2 \text{ chooses } C, \text{ pay-off must be } b_3^1. \end{cases}$$

It follows from (4.1) that P is the correct choice, hence,

$$b_2^1 > b_3^1. \tag{4.6}$$

By a similar set of arguments

$$\begin{aligned} \text{In } \mathcal{P}_2, \mathcal{P} \text{ chooses between } & \begin{cases} P, \text{ pay-off must be } b_2^2, \\ \text{or} \\ A, \text{ pay-off must be } b_3^2, \end{cases} \\ & b_2^2 > b_3^2 \end{aligned} \tag{4.7}$$

and

$$\begin{aligned} \text{In } \mathcal{P}_3, \mathcal{P} \text{ chooses between } & \begin{cases} P, \text{ pay-off must be } b_2^3, \\ \text{or} \\ A, \text{ pay-off must be } b_3^3, \end{cases} \\ & b_2^3 < b_3^3. \end{aligned} \tag{4.8}$$

At this point, it may be appropriate to sum up what we actually have found so far. The set of inequalities (4.9) presented in the box below must be satisfied to secure the NE (4.1)

$$\begin{aligned} a_2^1 \cdot p_1 + a_2^2 \cdot p_2 &> a_1^1 \cdot p_1 + a_1^2 \cdot p_2, \\ a_3^3 &> a_4^3, \\ b_2^1 &> b_3^1, \\ b_2^2 &> b_3^2, \\ b_2^3 &< b_3^3. \end{aligned} \tag{4.9}$$

Now, a reasonable continued analytic strategy may be to repeat the above calculations for a different NE - namely, the (EXPECTED I;OBSERVED) NE. If we apply the alternative (and more suitable) notation, this one may be expressed as; $(P|1, P|2, A|3; C|A, C|P)$.

We choose not to repeat these calculations in full, simply note that the inequality sign in Equation (4.4) must be reversed, while Equation (4.5) remains unchanged.

For Equations (4.6), (4.7) and (4.8), the change of NE from $(P|1, P|2, A|3; C|A, R|P)$ to $(P|1, P|2, A|3; C|A, C|P)$ leads to changes where all b_2^i must be replaced by $b_1^i \forall i = 1, 2, 3$.

Consequently, we can sum up our findings in this situation as

$$\begin{aligned} a_2^1 \cdot p_1 + a_2^2 \cdot p_2 &< a_1^1 \cdot p_1 + a_1^2 \cdot p_2, \\ a_3^3 &> a_4^3, \\ b_1^1 &> b_3^1, \\ b_1^2 &> b_3^2, \\ b_1^3 &< b_3^3. \end{aligned} \tag{4.10}$$

5. DISCUSSION, CONCLUSIONS AND FURTHER RESEARCH

5.1. Discussion and Conclusions

At this point, it may be fruitful to sum up our results so far. We have chosen to analyze two different NEs, the (EXPECTED I; DEFAULT) NE and the (EXPECTED I;OBSERVED) NE. The “boxed” inequalities (4.9) and (4.10) state necessary constraints for realizing each of these NEs respectively.

Simply put, we can interpret the (EXPECTED I; DEFAULT) NE as a game outcome in which adequate strictness is applied to patients, while the (EXPECTED I;OBSERVED) NE involves more than adequate strict treatment. Alternatively, and perhaps less normatively, we can interpret the two situations as a patient-dependent – as opposed to a patient-independent treatment scheme.

The reason why we are interested in comparing these two situations is an inherent hypothesis (to some extent confirmed by [20–23]) of real world existence of too much patient independent treatment in psychiatry. Given such a hypothesis, the constraints (4.9), (4.10) tell us how it might be possible to move from one (undesirable) situation to another (desirable) situation. In practice, such a change could involve various regulative means – here symbolized through the manipulation of model parameters, b 's, a 's and p 's.

If we move on comparing inequalities (4.9) and (4.10), we observe immediately that the second inequality ($a_3^3 > a_4^3$) is unchanged between the two NEs. The actual meaning of this inequality is that an aggressive patient (type 3) should be treated correctively (the therapist should choose C with greater utility than choosing R) if the patient signals aggression (patient choice A). In this setting, such an outcome means that any regulative action involving these two parameters will have no impact whatsoever on the two given NEs. From a game theory point of view, this result might be considered banal, but it still introduces a certain practical lesson. Certain infrastructural settings may not have any impact on relevant Nash Equilibria.

Another far more interesting observation to make in the comparison between (4.9) and (4.10) is a certain mathematical *separability*. By this, we refer to the fact that the structure of the equations within each box (4.9) as well as (4.10) indicate a clear separation between b 's on side and a 's and p 's on the other side. None of these inequalities bind all parameters together. This implies an easier regulative situation, than we might fear. For instance, the p 's in the model indicates the degree of uncertainty the therapist has related to the patient type. The structure of the inequalities tell us that if we (say) compare a situation with very limited information (p 's closer to $\frac{1}{3}$) of patient types with a situation with greater precision (p 's close to 0 or 1), it is only necessary to analyze a single inequality - namely the first one.

Alternatively, we can interpret this observation as a regulative separability between patient and therapist. Wishing to change parameters in order to achieve some desirable NE, we could change utilities for patient and therapist one by one, without a need to see them together. This is good, in the sense that it eases practical regulative measures.

Our final (and as we see it, most interesting result) is slightly more complex. In order to arrive at it, we need to make certain assumptions. Let us assume the following:

$$b_2^i = b_1^i + \epsilon, \epsilon > 0 \text{ but small.} \quad (5.1)$$

Before we explain the meaning of this assumption (5.1), note that it implies:

$$b_2^i > b_1^i \quad \forall i \in \{1, 2, 3\}. \quad (5.2)$$

Referring to Figure 2, it is straightforward to realize that (5.1) means that any patient type launching a passive (*P*) signal should be treated with recognizing (*R*) rather than Corrective (*C*) action. In our opinion, such an assumption cannot be considered controversial.

The next step in the argument is to assume that we are in the (undesirable) NE. That is, inequalities (4.10) are all satisfied.

Now, let us inspect the alternative NE – or inequalities (4.9) and start by examining inequalities 3 and 4. They are

$$\begin{aligned} b_2^1 &> b_3^1, \\ b_2^2 &> b_3^2. \end{aligned} \quad (5.3)$$

Given assumption (5.1) and its consequence (5.2), it should be relatively straightforward to realize that inequalities (5.3) also must be satisfied³.

Finally, we know that $b_1^3 < b_3^3$, but we also know that $b_2^3 > b_1^3$. Now, the ϵ comes in handy, because we can always find an ϵ small enough to secure that

$$b_2^3 < b_3^3.$$

Now, we have shown the following: Given our (reasonable) assumption in (5.1), all inequalities 2 up to and including 5 in (4.9) are already satisfied, and the only remaining inequality we need to regulate is the first one, which needs to be changed from

$$a_2^1 \cdot p_1 + a_2^2 \cdot p_2 < a_1^1 \cdot p_1 + a_1^2 \cdot p_2$$

to

$$a_2^1 \cdot p_1 + a_2^2 \cdot p_2 > a_1^1 \cdot p_1 + a_1^2 \cdot p_2$$

where the only difference is the direction of the inequality sign.

Apart from the fact that we have obtained a simple (numeric) expression defining the move from an undesirable NE to a desirable one, this result has another obvious consequence. If our aim is to move from a patient independent to a patient dependent treatment outcome, it makes no sense to regulate patient infrastructure. On the other hand, it makes all sense to regulate therapist infrastructure. This is an obvious consequence of the argument above, and may not be intuitively obvious. Hence, it makes for instance no sense to believe that improving conditions for patients in any way will have any impact on treatment strategies in the therapist's hand. On the contrary, what we have shown above, is that only therapist pay-off (or utility – *a*'s) combined with therapist knowledge and understanding of the patient (*p*'s) matter.

³If $b_2^1 > b_1^1$ and $b_1^1 > b_3^1$ then $b_2^1 > b_1^1 > b_3^1$ and $b_2^2 > b_3^2$.

5.2. Model limitations and suggestions for further research

Obviously, our results obtained and discussed in Subsection 5.1 are based on the model at hand. At best, this model is an extreme simplification of reality. Note for instance, that the only incomplete information bit is from the therapist to the patient. It would probably be far more realistic (possibly also more interesting) if a more “symmetric” asymmetry had been introduced. Say, the patient is also unsure on the therapist-type. In addition, our modeling framework limits itself to a one-shot treatment situation, also very different from reality. In the real world, a therapist and a patient, would meet more than once, possibly learning about each other during such a process. This would of course lead to a dynamic game of incomplete information, unfortunately not easy either to formulate or to analyze. Not to mention the fact that modern treatment involves more than a single patient/therapist relation, calling for a multi-player setting.

Still, we feel that this way of attacking the problem is both new and interesting. It opens up for a more analytic way of analyzing how such relations should be constructed as well as possibly regulated, in order to maintain and reach certain goals related to treatment quality. We do (obviously) not see this as an endpoint, rather a start of a different way of research within psychiatric nursing. As pointed out above, many directions exist when it comes to refining the game theoretic modeling framework. Another line of possible application could be a more empirical attack, where one tries to actually estimate all model parameters in order to explicitly find necessary regulative strategies (i.e., finding out which and how many various parameters must be changed in order to achieve certain goals.)

Let us finally point out, that in this work, we have focused on psychiatric care. As most readers might realize, this model concept is not limited to the psychiatric case. For instance, certain somatic diseases resemble this setting. Think about hypochondriacs or people who want a certain diagnosis. Hence, situations where the challenge for the therapist is to decide whether the patient is sick or not. Obviously a candidate for similar models.

REFERENCES

- [1] P. Anand, *Blame, game theory and economics policy: The cases of health and public finance*, J. Theor. Polit. **10** (1998), 111–123.
- [2] N. Bellomo and M. Delitala, *From the mathematical kinetic, and stochastic game theory to modelling mutations, onset, progression and immune competition of cancer cells*, Phys. Life Rev. **5** (2008), 183–206.
- [3] A. M. Colman, *Cooperation, psychological game theory, and limitations of rationality in social interaction*, Behav. Brain Sci. **26** (2003), 183–206.
- [4] A. M. Colman and J. C. Wilson, *Antisocial personality disorder: An evolutionary game theory analysis*, Legal Criminol. Psych. **2** (1997), 23–34.
- [5] S. B. Dowd, *Applied game theory for the hospital manager: Three case studies*, Health Care Manag. **23** (2004), 156–161.
- [6] J. Duxbury, *An evaluation of staff and patient views of and strategies employed to manage inpatient aggression and violence on one mental health unit: A pluralistic design*, J. Psychiatr. Mental Health Nurs. **9** (2002), 325–337.
- [7] E. Goffman, *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*, Doubleday & Company, New York, 1961.

- [8] J. C. Harsanyi, *Game theory and the analysis of international conflicts*, Austr. J. Polit. Hist. **11** (1965), 292–304.
- [9] G. Høyer, L. Kjellin, M. Engberg, R. Kaltiala-Heino, T. Nilstun, M. Sigurjónsdóttir and A. Syse, *Paternalism and autonomy: A presentation of a Nordic study on the use of coercion in the mental health care system*, Int. J. Law Psychiat. **25** (2002), 93–108.
- [10] D. Lee, *Game theory and neural basis of social decision making*, Nature Neurosci. **11** (2008), 404–409.
- [11] Y. Mansury, M. Diggory and T. Deisboeck, *Evolutionary game theory in an agent-based brain tumour model: Exploring the ‘genotype-phenotype’ link*, J. Theor. Biol. **238** (2006), 145–156.
- [12] H. L. I. Nijman, *A model of aggression in psychiatric hospitals*, Acta Psychiat. Scand. **2** (2002), 142–143.
- [13] E. Rasmusen, *Games & Information*, 3rd ed., Blackwell Publishers Inc., Maldes Massachusetts, USA, 2009.
- [14] T. C. Reluga, *Game theory of social distancing in response to an epidemic*, PLoS Comput. Biol. (2010), DOI: 10.1371/journal.pcbi.1000793.
- [15] J. E. Riggs, *Medical ethics, logic traps, and game theory: An illustrative tale of brain death*, J. Med. Ethics **30** (2004), 359–361.
- [16] A. E. Roth, *Game theory as a part of empirical economics*, Econ. J. **101** (1991), 107–114.
- [17] C. Tarrant and A. M. Coleman, *Continuity and trust in primary care: A qualitative study informed by game theory*, Ann. Fam. Med. **8** (2010), 440–446, 2010.
- [18] C. Tarrant, T. Stokes and A. M. Coleman, *Models of medical consultation: opportunities and limitations of a game theory perspective.*, Qual. Saf. Health Care **13** (2004), 461–466.
- [19] I. P. M. Tomlinson, *Game-theory models of interactions between tumour cells*, Eur. J. Cancer **33** (1997), 1495–1500.
- [20] S. Vatne, *Korrigere og anerkjenne. Sykepleieres rasjonale for grensesetting i en akuttpsykiatrisk behandlingspost*, Ph.D. thesis, University of Oslo, August 2003.
- [21] S. Vatne, *Limit setting in mental health: historical factors and suggestion as to its rationale*, J. Psychiatr. Mental Health Nurs. **13** (2006), 588–597.
- [22] S. Vatne, *To correct and to acknowledge: two simultaneous and conflicting perspectives of limit-setting in mental health nursing*, J. Psychiatr. Mental Health Nurs. **14** (2007), 41–48.
- [23] S. Vatne, *Korrigere og Anerkjennne*, 3rd. ed, Gyldendal Akademisk, Oslo, 2009.

Kjetil K. Haugen, Faculty of Logistics, Molde University College, Specialized University in Logistics, Britvegen 2, 6410 Molde, Norway
e-mail: kjetil.haugen@himolde.no

Solfrid Vatne, Faculty of Health Sciences and Social Care, Molde University College, Specialized University in Logistics, Britvegen 2, 6410 Molde, Norway
e-mail: solfrid.vatne@himolde.no

