D6.2 Final Concept Assessment Report

Deliverable ID:	D6.2
Dissemination Level:	PU
Project Acronym:	BEACON
Grant:	893100
Call:	H2020-SESAR-2019-2
Topic:	SESAR-ER4-08-2019
Consortium Coordinator:	University of Westminster
Edition date:	30 November 2022
Edition:	01.02.00
Template Edition:	02.00.05

Authoring & Approval

Authors of the document

Name / Beneficiary	Position / Title	Date
Andrea Gasparin / UNITS	Project member	28/10/2022
Gérald Gurtner / UoW	Project Coordinator	07/11/2022
Tatjana Bolic / UoW	Project member	31/10/2022

Reviewers internal to the project

Name / Beneficiary	Position / Title	Date
Benno Guenther / Salient	Project member	03/11/2022
Marie Carre / SWISS	Project member	03/11/2022
Nadine Pilon / EUROCONTROL	Project member	07/11/2022
Andrew Cook / UoW	Project member	07/11/2022
Graham Tanner / UoW	Project member	07/11/2022
Lorenzo Castelli / UNITS	Project member	07/11/2022

Reviewers external to the project

Name / Beneficiary	Position / Title	Date
N/A		

Approved for submission to the SJU By - Representatives of all beneficiaries involved in the project

Name / Beneficiary	Position / Title	Date
Gerald Gurtner / UoW	Project Coordinator	07/11/2022

Rejected By - Representatives of beneficiaries involved in the project

Name and/or Beneficiary	Position / Title	Date
N/A		

Document History

Edition	Date	Status	Name / Beneficiary	Justification
00.00.01	14/10/2022	Internal	Andrea Gasparin / UNITS	Draft for SJU
00.00.02	31/10/2022	Internal	Gerald Gurtner / UoW	Internal version
01.00.00	02/11/2022	Internal	Gerald Gurtner / UoW	Internal review
01.01.00	07/11/2022	Submitted	Gérald Gurtner / UoW	For SJU submission

01.02.00	30/11/2022	Submitted	Gérald Gurtner / UoW	Second submission after SJU feedback

BEACON

BEHAVIOURAL ECONOMICS FOR ATM CONCEPTS

This deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 893100 under European Union's Horizon 2020 research and innovation programme.

Abstract

In this deliverable we present the final concept assessment analysis. We review the stakeholders' feedback collected during the human-in-the-loop simulation (HITL) activity, concentrating on the benefits as well as the weaknesses of the proposed mechanisms, and with a particular focus on their feasibility in terms of real scenario implementation. We will compare the HITL considerations with the various simulations drawing final conclusions and recommendations for future research and how this could take advantage of the lessons learned during the project.

Table of Contents

List of Figures

1 Introduction

BEACON is interested in estimating the impact of new mechanisms for the resolution of hotspots. Indeed, when a piece of airspace is predicted to be over capacity, a.k.a a 'hotspot', some flights need to be prioritised over others in order to keep the traffic under the maximum capacity provision. The mechanism by which the flights are prioritised is usually the First-Planned First-Served (FPFS), which minimises the total delay incurred by the flights. In order to take into account other aspects, such as airspace user (AU) costs, BEACON has tested several mechanisms (see D3.1 [1] and D4.2 [2] for more details):

- ISTOP (Inter-airline Slot Trading Offer Provider): allows airlines to swap pairs of flights between themselves.
- Credit mechanism: allows airlines to spend virtual credits to declare flight costs.
- Auction: allows airlines to bid for slots using virtual credits.

These mechanisms have been explored in BEACON in D4.1 [3], D4.2 [2], D5.1 [4], and D5.2 [5] with the help of human-in-the-loop and fast-time simulations using behavioural economics to assess realistic decision-making processes from humans involved in the mechanisms.

In this deliverable we provide an overall assessment of the BEACON concept. More specifically, we discuss the potential implementation of the above mechanisms by discussing several aspects which are not directly linked to their efficiency, as estimated by the different models using during the course of the project.

We are interested in discussing several aspects. First, we would like to come back to several issues that were identified during the course of the project linked to the design of the mechanisms themselves. While we discuss some of the choices made by the consortium, we are interested in providing some future general guidance for those aiming at designing such mechanisms. These issues are related to using virtual credits vs money, a centralised optimiser vs a decentralised trading mechanism, gaming effects, and of course behavioural ones.

Second, we are interested in highlighting some of the limitations of the mechanisms presented in BEACON, while, once again, trying to draw some general conclusions to inform future research. We will discuss some potential barriers to implementation, from different points of view, or serious issues that were simplified or purposely ignored in BEACON but should be addressed in any future research tackling the same subjects.

Finally, we are interested in highlighting the methodological improvements that could be used in future for the **estimation** of the mechanisms' impact (from BEACON, or otherwise). These aspects are more theoretical and linked to the choice of the framework used for the modelling process, in particular with respect to gaming effects and behavioural ones.

With this deliverable, we hope future research will be made aware of some of the key issues and to find a line of work that would build on the work performed in BEACON, even when designing brand new mechanisms for hotspot resolution processes.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

The deliverable, a collection of discussions on the key aspects highlighted above, is organised by using three points of view: the airspace user in Section 2, the air traffic manager in Section 3, and the modeller in Section 4. We provide our consolidated conclusions and recommendations in Section 5.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

2 Airspace user's point of view

In this section, we explore how the mechanisms explored in BEACON may be received by the AUs. Indeed, regardless of whether our estimation of the mechanisms' impact is accurate (an issue that we explore in Section 4) and whether the impact is positive or not (an issue that we explore in Section 3), the question of whether the mechanisms are suited to the airline's workflow needs to be addressed. Human factors obviously play a major role in this, and, while being slightly out of the scope of the BEACON, we use the experience of the human-in-the-loop simulations to draw some conclusions. The issue of the adoption of the mechanisms by the AUs is however bigger than just the human factors problem, and we discuss other facets in the following subsections.

2.1 Human-in-the-loop simulations

Within the WP5 human-in-the-loop simulations (HITL)[1], real-time simulations were performed. The primary goal of these experiments was to compare the behaviour of real dispatchers against their modelled counterparts. In particular, this task focused on estimating the importance of behaviours deviating from 'rationality'[2].

To carry out this exercise the development of a human-machine-interface (HMI) was necessary, and the adaptation of the Mercury simulator[3] to enable the interaction between the participants and the simulator. The HMI enabled the participants to play the dispatcher role which could have been otherwise played by an artificial intelligence subroutine in the simulator.

The HMI was presented as a table that contained a series of data fields (columns) for each affected flight (row) with information related to schedules and regulations. In addition, the HMI contained a series of columns where the participant could input/alter the values for the flight, in order to carry out prioritisations for selected flights. Figure 1 and Figure 2 show the HMI for the Credits and ISTOP mechanisms, respectively. Below the ISTOP mechanism a line showing the costs for a flight in relation to the slot in the regulation is shown.

[1] This part comes directly from D5.2. We reproduced here for convenience and because we believe that the explanations on how the human-in-the-loop simulations were performed shed a light on the following arguments.

[2] As defined in previous deliverable (and in the economic/game theory field in general), rationality is defined as the course of action that would lead to the best outcome for a given agent, given a situation and the level of information accessible to the agent.

[3] As reminder (it is explained more in details in other deliverables, see D3.1, D5.1, and D5.2), the Mercury simulator is a model developed over several years to compute advanced indicators on the air (and later ground) transportation system, such as the distribution of delays for passengers, flights, etc. It uses a strong agent-based paradigm, particularly suited to heterogeneous systems.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

Figure 1. HMI for Credit mechanism.

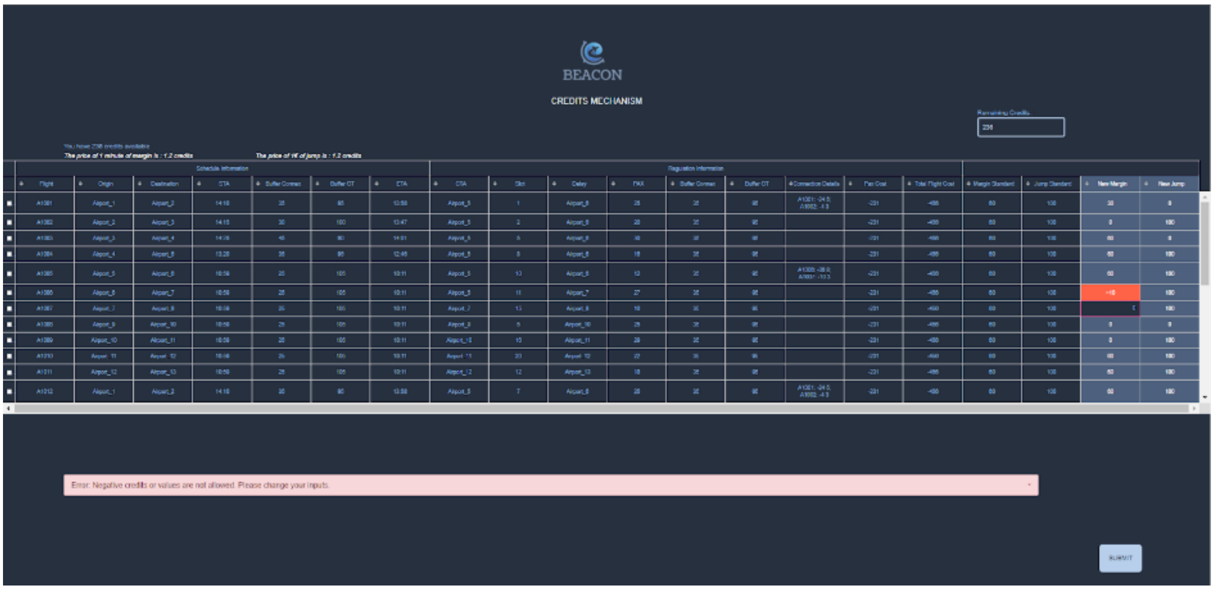

Figure 2. HMI for UDPP+ISTOP mechanisms with cost across slots information.

When the participant finished editing the cells with their input on how to resolve a hotspot, the 'submit' button was pressed and the information sent to the simulator. When ready, the results were presented to the participant through another HMI view.

Exercise sessions were scheduled to last between two and two and a half hours. The consortium organised eight exercise sessions with seven participants, as the first session had to be interrupted (due to incompatibility with the participant's browser – Safari). Six of the exercises were successful, in terms of collecting and exchanging data with the simulator. One exercise had to be dropped because of technical incompatibilities with the participant's browser (Safari) and the last one could never be scheduled.

The exercise was composed of an introduction, and two sessions with BEACON mechanisms. Each session started with ISTOP or the Credit mechanism in a random manner. The BEACON team was in contact with the participant throughout the exercise, to troubleshoot any glitches, or answer questions on the HMI or mechanisms' characteristics. During the exercise, the team members were collecting feedback from the conversations. At the end of the exercise, the participants were asked to respond (verbally) to a list of qualitative questions, and to fill in the system usability survey (SUS) for the two mechanisms.

The qualitative feedback questions were:

- Did you feel comfortable exchanging slots with other airlines?
- Was it an issue that you could not see the results for other airlines? If so, in what way?
- Did you make your choices purely on cost minimisation? If 'no', please summarise other factors driving your decisions.
- When making your choices, what were the main objectives you had in mind?
- What did you think about the ISTOP mechanism? Were you expecting these results?
- What did you think about the credits mechanism? Were you expecting these results?

None of the participants had issues with exchanging slots with other airlines. Furthermore, they did not mind not having access to the results for other airlines, as they were focusing on the best solutions they could provide for their own airline.

Some participants made their choices purely using the cost information presented via the HMI, while others took into account other factors, as they would in their daily job. These ranged from trying to obtain the best on-time performance for their flights, over favouring flights with connections, minimising the minutes of delay for passenger, to taking care of higher passenger categories (e.g. business passengers), sometimes assuming that a change of aircraft would be possible for downstream rotations. When taking only the costs into account, the participants tried to obtain the slot just before a big jump in the costs.

The opinions on the ISTOP mechanism span a spectrum. Some participants declared that it was not intuitive and could not understand the consequences of submitted choices. Other participants declared that the mechanism was understandable, even if it required some learning for familiarisation. The possibility of implementing their own scales for margins and jumps was appreciated. These participants noted the improvement in the costs using the ISTOP.

The opinions on the Credit mechanism were also divided. For some participants it was easier as this mechanism offered standard values. In case of doubt (in a sense of what to do with a particular flight), the standard values were left in place. However, other participants questioned the meaning of 'standard values' as it was not clear to them how those were set. What was appreciated in this mechanism was that there was a clear trade-off in setting up jumps and margins – for a better position, some credits needed to be spent, and for accepting a worse one, credits were gained. Thus, in one participant's words, "there is a need to give up on some flights' time to build credits, to use it to protect the most valuable flights.” The time of creation and roll over of credits was discussed, as something that would be useful to know for the use of the mechanism.

Furthermore, the participants were asked to complete a questionnaire to collect system usability score (SUS) scores on both mechanisms.

The average score for the Credit mechanism is 65, while it is 70 for ISTOP. Apparently, usability of ISTOP is slightly better. However, with just six respondents, it is not possible to test the scores for statistical significance.

Lessons learned regarding the organisation and running of HITL exercises:

- The cost information for flights (as a function of the delay) was much appreciated by participants.
- Clear indication of cost savings through the choices is needed by participants. The HMI did show the initial regulated costs and costs after the submittal of choices. However, the visualisation of this information could have been better. The information was just displayed, so the participants should have gone through each flight and checked, remembering what they chose for each flight. We appreciate that this is not a very good indication, and that it could be improved.
- When limited time with participants is available, it is very important to have an efficient introductory session explaining the exercise objective, mechanisms and the HMI details. The consortium provided this briefing. However, we found that the same briefing did not work the same with all the participants, as for some all was rather clear, while for some others the learning curve was much slower.
- Need to take into account where the participants are connecting from. For example, most were connecting from their work computers, that had varied degrees of connection security in place. In some instances that delayed, or made the exercise impossible, as the HMI was designed with the goal of collecting the information, but not to be at a level that can be certified as secure connection by more stringent connection security filters.
- Need to take into account the browsers the participants would use. Most of the participants used Google Chrome or Microsoft Edge, with which HMI worked correctly. However, the setup did not work properly with Safari.

2.2 Create trust in the system

As a general consideration, arising from the various project stages (advisory board, HITLs and stakeholders’ feedback), one of the main hurdles that any potential new mechanism for ATFM slot assignment is that the policymaker should not only find efficient and equitable procedures and algorithms, but also design a strategy to ensure a wide consensus and trust among the stakeholders. In fact, if on one hand it is naturally simple for an AU to agree on a delay reduction for any of its flights, on the other hand, it is easy to consider it much riskier to accept a further delay deterioration, in particular if the benefits of such are not guaranteed due to the uncertain nature of the network, or even worse if those benefits are expected to materialise only in the long term. For example, all mechanisms tested in BEACON showed promising results when applied on a single regulation, but before considering their implementation in the real scenario, two additional stages are in our opinion required:

- to study the behaviour on a variety of data-driven test cases including the widest set of airports as possible and extend the analysis on the long-term effects;
- if also these aforementioned experiments show satisfactory results, a validation phase consisting of multiple *human-in-the-loop simulations* should be conducted, in order to build and consolidate the confidence of the stakeholders in the mechanisms proposed.

2.3 Cost models dependency

As already pointed out in D6.1 [6], Section 3.1.2, margins and priorities may not be enough to approximate real airline cost functions[4]. Indeed, in D6.1, we already mentioned that real cost functions can be particularly complex, because of missed connections, crew, etc. We pointed out then that even "flight margins", as conceptualised in UDPP and used in BEACON may represent poor approximation of the cost function and thus do not represent the full complexity of the operational data for airlines. This aspect is of particular relevance, as it has been shown by the simulation results. A poor approximation of the delay cost models might dramatically deteriorate the cost efficiency of the mechanism, resulting in a poor exploitation of such mechanisms, or even an economic loss for the AUs. This ultimately means that a key role in the explored mechanisms is played by the delay cost models approximations. In fact, the simulations showed the high sensitivity of almost all the mechanisms with respect to cost model simplifications. This outcome suggests that future research should put a particular focus on this topic as regardless of the benefits of the solutions proposed, it has been clearly shown that poor assumptions or approximation of the cost models may lead to poor efficiency of any mechanism.

2.4 Money vs credits

A similar consideration holds for the problem of money into credit conversion. As already mentioned in D6.1 [6], Section 3.2.3, monetary mechanisms are unlikely to be accepted by the users. Indeed, we have shown in D6.1 that virtual credits are a requirement from airlines for any trading mechanisms. First, because airlines have already paid for the ATFM service and consider that they should not pay again for it, and second because the operational side of the airlines would not deal with money directly. We pointed out then that, although credits would work in practice fairly similarly to a monetary-based mechanism, several choices have to be made concerning the use of credits (like how credits are created), which may impact the mechanism and need to be explored further.

For this reason, it has been explored how to develop and improve credit mechanisms. However, if cost approximation is a hurdle, things become even harder when defining an equivalence between credit and costs. The credit mechanism developed in BEACON showed promising results despite the previously mentioned limitations, but it also showed a significant sensitivity to the kind of approximation performed to compute, first, the cost model surrogate and, second, the credit conversion. Nevertheless, it must be remarked that when compared to the UDPP and ISTOP models, the flexible credits have a huge potential advantage which is the possibility to extend the scope on multiple regulations, an aspect which will be discussed in more detail in Section 3.3.

3 Air traffic manager's point of view

In this section we explore several issues related to the implementation of the mechanisms from the air traffic manager's point of view, highlighting the barriers to implementation arising from the interaction of these with other processes, the need to define adequate levels of automation, or the lack of performance of new mechanisms with respect to the current UDPP.

3.1 Compromise between centralised and decentralised mechanisms

As introduced in D6.1 [6], uncertainty plays an essential role in this context. From the air traffic manager's perspective, rescheduling a flight (slot allocation/swapping) might have knock-on effects on several operations. For example, for the airports each change of plan may require a re-allocation of resources (D6.1, Section 3.2.1) where instead, at the network level, a change of slot of a flight in a given regulation might lead to the saturation of another network resource, possibly causing additional disruptions (D6.1, Section 3.2.2).

In order to increase airlines' flexibility and involve them directly in the decision-making process it is required, and desirable, to introduce a certain degree of decentralisation, meaning that the NM delegates some decisions to the airlines[5]: from the purely centralised mechanism where the NM solves the regulation without feedback from the stakeholders (like FPFS) to the purely decentralised activities where stakeholders trade the necessary resources they lack (e.g. slots in the context of ATFM regulations).

However, due to operational requirements from different stakeholders and their conflicting preferences and objectives, it is clear that not all decisions can be delegated and even within those decisions which might be decentralised. In other words, one needs a 'conflict resolution mechanism' when requirements, or preferences, from different stakeholders are not compatible, or a limited selection of options for the users in order to request changes that satisfy operational constraints and are satisfactory for all stakeholders. Thus, a compromise between decentralised and centralised mechanisms is likely to represent the most reasonable approach to accommodate AUs' needs for flexibility and the ATM necessity of ensuring a certain level of efficiency and stability of the system.

In this respect, all mechanisms tested in the BEACON simulations, UDPP, UDPP+ISTOP and Flexible Credits are consistent with this idea. In fact, they all define an initial decentralised phase in which all AUs allowed state some preferences for their flights, either via prioritisation, bids for slots or cost surrogate declarations. In the second phase, a centralised optimisation takes place, using UDPP merge, global optimisation (CM), swap offers (ISTOP), or realised slot acquisition (Auction). In this phase, the

[5] In contrast with the application of the FPFS rule in particular, where the optimisation is done without any feedback from the airlines.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

central planner computes the actual final allocation, potentially taking into account operational constraints and preferences in a harmonious manner.

This mixed procedure, with some degree on decentralisation leading to a centralised optimisation, has several benefits. First, operational requirements and preferences may be communicated in a standard and comprehensive way to the central planner, contrary to a pure centralised version. Second, the central planner is in charge of defining what is “best” for the system, on the contrary of the pure decentralised version. Indeed, focusing only on economic performance (saved cost) as it would be the case for a trading mechanisms may lead to unwanted side effects, while keeping for instance some degree of equity in the mechanism may be desirable, and only possible with the presence of a central optimiser.

This mixed approach also has also some serious limitations. First, giving all power to a central planner to define the ‘best’ situation for the system may not be attractive to airspace users in particular. Even if the rules are to be chosen by consensus, it might be hard to reach one. We illustrated for instance in D5.2 how several different indicators of equity may obtain in different mechanisms, and different airlines may have different preferences. Second, even if the central planner knows ***what*** to optimise (saved cost + plus equity for instance), it may face difficulties getting the information needed for the optimisation. Indeed, as we have seen in D5.1 and D5.2, errors (intentional or not, due to approximations, gaming effects and behavioural ones) in the cost declaration may lead to very poor performance for the subsequent optimisation.

3.2 Need for automation

As already arose in the workshop and discussed in D6.1 [6], Section 3.3.1, one of the important issues in the development of ATFM slot allocation solutions is automation. This is clear in particular at the network manager level, as a solution which takes into account the AUs’ requests, and other operational constraints (for instance taking airport constraints into account) needs to solve a complex optimisation problem, a task which is impossible to perform without computer assistance.

This last requirement is respected by the current default mechanism (FPFS) meaning that the centralised phase is performed in a nearly fully automated manner[6]. On the other hand, the current mechanism for slot swapping is completely manual and relies on bilateral agreement between airlines. However, as confirmed by the workshop discussion, during the human-in-the-loop exercise and by the simulations, the slot assignment mechanisms must include a high level of automation also in the decentralised phase.

There are several reasons behind this conclusion: the first is the complexity of the task, for an AU dispatcher, to keep a comprehensive picture of all flights’ delay costs, accurate enough to ensure the efficient exploitation of any slot allocation mechanism. This becomes even harder if we consider that these mechanisms are meant to be deployed at a tactical level, on the day of the operation, which

[6] The Network Impact Assessment is computerised but launched manually by the Network.

implies a limited time to analyse the situation, activate and manage the decision process, with the overarching issue of most AUs being only partly aware of their tactical costs. Toward this purpose, several lines of work could be used in the future to provide solutions to this problem.

Indeed, in order to run the simulations, the decision-making process of an AU agent needed to be represented by an algorithm. This leads to the definition of different optimisation algorithms that in the model are meant to provide the best solutions for the AUs. For UDPP an integer programming algorithm has been implemented which is able to support all the airlines to compute their respective best prioritisation for their flights. For ISTOP and Flexible Credits, an efficient custom model approximation algorithm has been developed. We believe these improvements might represent an essential contribution for the development of the automation of the mechanisms. However, we want to highlight that these algorithms are not meant to entirely replace humans, instead they simply aim to support the airlines decision process and they do not preclude human intervention (being conducted for example by tuning the algorithm's parameters or simply by manually modifying the algorithm's outcome).

3.3 Beyond the single regulation concept

The results obtained by the simulations, not just in terms of cost reduction but also in terms of equity, show that to aim at a substantial improvement, it is necessary to go beyond the single regulation slot assignment problem and to consider the greater picture. This means to consider the entire network, possibly for longer period of time.

As already mentioned, UDPP provides significant results for airlines with a sufficient level of flexibility within a single regulation, *id est* more than 4 flights, but there is very little chance to improve the situations of those airlines who do not have such flexibility. Under this paradigm, ISTOP provides some additional gain of flexibility and consequently cost reduction which can certainly be considered as significant. However, its limitations, as for UDPP, lies in the fact that they are meant to operate on a single regulation and therefore one cannot expect that also further development will lead to any substantial improvement. It might be difficult to extend ISTOP to multiple regulations, due to the fact that it is based on the idea of slot exchange which requires to have defined a current slot assignment configuration in order to compute the slot swap offers. As such, it is not applicable in future regulations.

Conversely, the Credit Mechanism can be naturally adapted to take into account as many regulations as wanted without location or time restrictions. The Credit Mechanism is based on the 'give and take' idea, which also represents the way in which equity is implemented and is based on the intuition that airlines might find it convenient to accept a delay increase for certain flights if in exchange they obtain a delay reduction for others. The difference with ISTOP is that the credit mechanism potentially includes the possibility for an airline to exploit credits among different regulations. This means that, for example, if in a particular regulation an airline has only one flight with low priority, it might decide to accept additional delay, and gain some flexibility to use in other regulations, or maybe save them, i.e. planning to spend them in the future. This kind of mechanism represents for low volume operators in constraints (LVOCs) an opportunity to gain flexibility and ultimately increase their chances to benefit from a slot assignment procedure different to the current FPFS one. If on first inspection this aspect might not appear particularly relevant, data clearly show (like in D5.2) that LVOCs represent the majority of the airspace users involved in airport regulations, so in our opinion one of the most important conclusions of the project is that slot assignment mechanisms should be designed and studied to operate and provide benefits in different locations and at different times for the same airline.

3.4 Robustness of UDPP

Results of the simulations have proven the efficiency and the robustness of the UDPP mechanism[7]. Indeed, as showed in D5.1 and D5.2, the efficiency of UDPP is high compared to the other potential mechanisms*, **including the pure theoretical mechanism that would yield the best results from the cost point of view***. This represents an interesting assessment which emphasises the fact that by expanding the flexibility to allow each airline to optimise only its own resources already provides effective results. In addition, due to the UDPP structure, no significant delay deteriorations for other AUs is guaranteed, i.e. no flight will be delayed due to the decision of another airline. As expected, airlines with a great number of flights within a regulation are the ones who benefit the most from UDPP as they can efficiently exploit all UDPP features to (often) significantly reduce their delay cost (when compared to the FPFS assignment). It has to be stressed, that low volume operators in constraints (LVOCs), as already shown in several works struggle to benefit from UDPP, due to their lack of flexibility. Furthermore, the comparison within theoretical bounds, the 'GLOBAL'[8] and the 'NNBOUND'[9] mechanisms, show that there is still potentially a good portion of reduction that is not captured by UDPP and that the additional cost reduction provided by the other mechanisms actually do not fill the gap with the bound models and leave a great room for further improvement beyond UDPP.

[7] Here and in most BEACON deliverables, we designate the combination of Flight prioritisation and Selective Flight Protection (SFP) as the 'standard' UDPP. More details about this choice can be found in D3.1.

[8] The GLOBAL mechanism, defined in D5.1 and other deliverables, finds the flight/slot allocation that minimises the total cost across airlines.

[9] The NNBOUND, also defined in D5.1, is the GLOBAL mechanism with the added constraint that no airline can lose from the final allocation, i.e. that their cost cannot increase with respect to the FPFS allocation.

4 Modeller’s point of view

In this section we highlight some issues and make some considerations related to the modelling process used in BEACON. Indeed, in order to compute the indicators related to the impact of the mechanisms in different settings, we had to tackle several theoretical issues. The problem of slot exchange is not obvious for several reasons, including gaming effects and of course behavioural biases. The former occurs very often when agents are in competition for some resources, like a hotspot, and try to find the best outcome for themselves. Behavioural biases are usually apparent when a human is involved in a decision, especially when they need to estimate future gains in an uncertain environment. In the following we discuss some of the choices made during the project, some of the hypotheses, and we try to draw a plan for future research to fix some of the shortcomings of the present study.

4.1 Gaming effects

As shown in D5.2 [5], gaming effects are impossible to ignore when multiple actors interact with each other in order to get the best outcome for themselves out of a ‘collaborative’ mechanism. Gaming effects can be expected as soon as airlines are in competition for their slots and rules allow them to make decisions that may improve their situation (here, the final slots for their flights). This happens naturally in several mechanisms (ISTOP, CM, etc) when we are considering more rational agents than the naïve, ‘honest’ ones.

4.1.1 Rationality and gaming theory

The modelling framework for rationality that we presented in D5.1 [4] was very simplified in terms of the decision-making process of the airlines. Indeed, we considered a sort of mean-field approximation where agents were using the approximated optimal strategy corresponding to other players being honest. More generally, we have not explored the game within the context of game theory, where one needs to explore mixed strategies, rewards associated with strategies, etc., in order to have a better estimation of the decisions of the agents and thus the actual values of the indicators. We believe that such an analysis needs to be carried out when a specific mechanism comes closer to prototyping. Hence, a future line of research for this work is to have a better theoretical basis exploring game theory to its fullest.

4.1.2 Reinforcement learning

Even with a game simplified to fit a framework with discrete strategies, as would likely be required with game theory, the solution to the problem (i.e. which strategies would likely be followed by which airlines) would require numerical simulations. A good framework for this is reinforcement learning (RL), a way of finding the best decision to follow for an airline given its state (for instance, the number of credits it has, the number of flights in the regulation etc.). In the context of discrete choice games, it can be viewed as a numerical way of estimating the reward matrix required to solve the game in the context of game theory.

Reinforcement learning is often used as an optimisation process, using a learning agent to learn a particular task (e.g. playing video games). Lately, multi-agent reinforcement learning has been used to have several agents solve a particular task (e.g. deconflicting aircraft), usually with a good degree of collaboration. This last framework opens the possibility to study games where a high number of agents compete towards similar objectives, as is the case in the BEACON mechanisms. For this reason, we plan to further develop this line of research by using this framework, which will lead *in fine* to a better estimation of the impact of the mechanisms.

Note also that RL is particularly powerful when a time dimension is present, i.e. when the state of the agent depends on their past decisions. This would not be the case for all mechanisms, as highlighted in D5.1 [4] and D6.1 [6]. While it would be the case for CM and the Auction mechanism, because airlines carry their credits from one regulation to the other, it is not the case for ISTOP. In this case, the mathematical formulation is much simpler and may require only a simpler version of RL, which in general is very heavy, computationally speaking.

4.1.3 Gaming in the context of UDPP mechanisms

While we strongly believe that further research is needed with game theory and RL if any 'collaborative' mechanism is to be implemented, we also note that it might be difficult for airlines to actually game the mechanism because there is a strong incentive not to, which is their complexity.

Indeed, gaming strategies in some cases are obvious, for instance in the benchmark GLOBAL[10]. However, we note that strategies become a lot more complex in the context of the main BEACON candidates, in particular the auction and the CM[11].

As shown in D5.1, it is for instance not obvious for an airline to balance the credits spent during the regulation with the expected gains. Or in other words, to estimate the expected gains in cost when spending a certain amount of credits. As noted in D5.1, because of issues linked to the approximation process at the heart of CM, more parameters that the simple couple margin/jump are required, which increases the complexity of the decision-making process. As a consequence, we may expect airlines to choose the simplest solution, giving their best approximation for their costs, a.k.a. 'honest' behaviour.

This is even more challenging with the auction mechanism as an 'honest' behaviour is not easily definable. As a consequence, optimal strategies may be difficult to compute and rather counterintuitive[12]. In this context we do not see how an airline could try to game the system[13], and in fact the main challenge here is linked to the usability of the mechanism, linked to the problem of automation as explained in Section 3.2. In fact, we note that the introduction of automation may modify the conclusions above, rendering gaming easy enough so that it has to be taken into account.

---

[10] See footnote 88.

[11] Gaming in the context of ISTOP may also exist, but we believe that it would be much milder. We plan to use the models developed in BEACON to explore this question in a subsequent publication.

[12] For instance, we proved that an optimal strategy for airlines would be to put a very high bid on their first slot, and a very low one for their second slot, and that was only in a very simplified setup. The analytical work that leads to this result is presented in annex in D5.1 but the corresponding results will be published at a later stage.

[13] Note that in some cases it might be easier to try to 'disbenefit' a competitor rather than finding the absolute best situation for oneself. This kind of behaviour, as well as any other behaviour relying on simplified heuristics, should not be discarded too early and should be studied further.

4.2 Behavioural effects

One of the goals of BEACON was to explore the potential impact of behavioural biases that may appear when humans are faced with making decisions involving uncertainty and long-term planning. Our implementation of these biases in the models is fairly simple because it is based on a highly simplified definition of rationality. The results show that these biases have an impact on the efficiency of the mechanisms (see D5.1 and D5.2), comparable or greater to a potential gaming effect, i.e. decreasing the efficiency by 3-8% percentage points.

The model predicting the decisions of the airlines is also approximated in terms of the parameters that we used for the prospect function. Indeed, the prospect function is the function that defines the expected gain in terms of *utility* as a function of the real gain[14]. While these should ideally have been calibrated using the questionnaire that we published and for which we did not receive enough responses, we decided to use values that are commonly viewed as a good first approximations by the community[15]. The degree to which this choice influences the results is unknown and should be explored further.

It is important to remember that the presence of behavioural biases is conditioned on the exact role played by the human(s) in the process and, more generally, to any contextual input (for instance the way the human-machine interface (HMI) is built). It is thus misleading to have a very accurate estimation of the impact of these biases before a prototype is defined in detail. However, it is clear that we showed their potential impact in this context.

Overall, we strongly suggest that behavioural biases as well as the impact of context dependent human behaviour and decision making should not be ignored in the process of designing a process to solve a hotspot but taken carefully into consideration. Considerations of the role of humans, the degree of automation and the type of biases that could appear should be made. We also suggest that more research is required to investigate the worst effects that can arise from these biases, as well as effects that we did not study in BEACON but that are likely to play a role, such as the endowment effect[16]. Finally, we note a need for a fundamental line of research bridging the gap between reinforcement learning/game theory and behavioural economics so that they can be taken into account at the same time harmoniously.

[14] As a reminder, the utility and prospect theory assume that agents maximise the expected value of an abstract quantity called 'utility' or 'prospect', instead of their real gain.

[15] Obtained from a literature review, described in D4.1.

[16] The endowment effect is a bias under which an agent has a tendency to retain a certain object or advantage (for instance, a slot) instead of choosing another one that would be objectively better (for instance, another slot giving a smaller cost for a flight).

5 Conclusions

BEACON explored several new mechanisms designed to overcome the shortcomings of the current ones, referred to as UDPP in the BEACON context, with the help of behavioural economics, a theoretical framework developed to help predicting humans' decisions better.

A first important conclusion from the project is related to the efficiency of the mechanisms. Indeed, as highlighted in D5.1 [4], D5.2 [5], and in Section 3.4, the 'basic' UDPP tested here (consisting mostly of priority attributes) displays an efficiency already very high with respect to the maximum possible gains (from the total cost point of view). While this is good news for UDPP itself, this means that any new mechanisms for hotspot resolutions need to be carefully designed to make any additional gain. In fact, as highlighted in D5.1 and again in Section 2.3, any error of the AUs' costs will lead to poor performance for new mechanisms, at least with respect to UDPP. While the efficiency of UDPP might also be impacted by this issue, the multipartite nature of any 'collaborative' (or rather competitive) mechanism may prevent obtaining gains good enough to make difference.

Furthermore, while some errors may be linked to poor cost computation capabilities, others may arise directly from gaming effects. We have shown that agents seeking their own profit over anything else (a.k.a rationality) can severely impair a mechanism (by up to 10 percentage points in efficiency for instance). Furthermore, behavioural effects may distort the players' decisions to such an extent that the gains made on top of UDPP can be lost. This is a crucial point, because gaming effects are normally absent from UDPP and there is good reason to think that behavioural ones would be very minor. As a consequence, we recommend having any new mechanisms tested at the design stage taking into account both gaming and behavioural effects. Any mechanism that exhibits only marginal improvements over UDPP should be abandoned or further improved.

This leads to our second conclusion and recommendation. Indeed, by implementing rational and bounded agents, the consortium found some serious theoretical issues, that cannot be overcome only by using more powerful machines. In fact, we identified the definition of rationality and its solvability as being at the heart of a proper estimation of the mechanism's impact. Reinforcement learning is thought to be a good framework to have a better estimation, even though the solvability of the model would still be an open question (but this time related to the computational power). Thus, we suggest that any future research should consider this method in order to obtain a good estimation of the impact of a mechanism featuring competitive behaviour for slot allocation.

Moving on to slightly more practical issues, we would like to insist on a crucial aspect, in the form of a paradox: while the ultimate objective of an airline may be to increase its profit[17], airlines may not want to mix operational aspects too closely with financial aspects. Indeed, flight dispatchers for instance may have only a high-level knowledge of the cost associated with their decisions (e.g. using the number of missed connections), which leads airlines to want to decrease their cost in real money (which operational speaking translates into reduced delay in operations, smoother operations, better passenger experience, decreased reactionary impact etc) while preferring non-monetary. This difficult

[17] In the BEACON context, this means decreasing its costs, either directly operationally (e.g. related to a curfew) or related to brand image, for instance.

aspect can be partially solved using virtual credits[18], but may be part of a bigger change of paradigm in Europe, together with 4D trajectories for instance, where 'operational contracts' are part of the day-to-day life of an airline. We thus suggest future research to deal with this aspect, both by probing the airlines on their current practices and by imagining their future ones.

This is linked to our fourth conclusion, concerning automation. Indeed, due to the complexity of solving a multi-partite mechanism where airlines compete for the best slots, it is tempting to just imagine a magic button that the airlines can press in order to have the best deal. However, it should be clear that, contrary to a local optimisation like UDPP, **such a button cannot exist** in general. Indeed, as highlighted at various stages of the project, the final allocation of an airline for its flights depends on the decisions made by all the other airlines. This is in sharp contrast with the simple flight prioritisation process from UDPP where the final allocation for an airline depends only on its own decision (the priorities on its flights)[19]. What is the meaning of having a full automated system in this case? Unless there is dominating strategy for the airlines[20], there is no easy way to have one.

This does not mean however, that automation will not play a role. Pre-computation of possible outcomes, for instance, may be used, as well as straightforward cost computations[21] from the airline. However, a human decision seems to be likely at some point in the process. Hence, we suggest that any further research reflects on the level of automation that could be expected from the system and actors, given the available techniques and data flows, at the design stage. It may be that some mechanisms (like the primary auction) may require a degree of automation that is out of reach of the current system, or any future one. The mechanism usability is thus intimately linked to automation.

Another important aspect for adoption is the perceived efficiency and fairness of the mechanisms, related to the trust in the system. Indeed, we have explained that, in order to reach higher levels of efficiencies, we need to allow airlines to lose locally in time and space in order to gain in the longer run. This means that airlines may lose from a hotspot resolution, and may even lose several times in a row. Any new mechanisms thus need to prove to be beneficial in the longer-run, and replace a 'nobody ever loses' paradigm by an 'everybody gains in the long-run' one. This can be done by fast-time simulations, and also extensive human-in-the-loop simulations showing the benefits of the mechanisms. Given the stochasticity of the system, this is a hard task, since humans may struggle to estimate properly the long-term benefits. In fact, we are expecting humans to display some strong behavioural biases when faced with the adoption of the mechanism, including of course the status quo bias. We thus suggest any future research to estimate *ex ante* which kind of proof would be acceptable to airlines for adoption and to communicate skilfully on the expected gains of the mechanism(s).

---

[18] Note that credits are linked to another issue, the reluctance of airlines to communicate their costs if there is a chance that competitors have access to them. This can be solved via a trustworthy central authority (like the network manager) or via some blockchain technology. In any case, our conclusion from BEACON on the subject is that this issue can be fairly easily overcome and should not constitute a major focus in any future research dealing with the design of a mechanism.

[19] Note that there is still the caveat of the Selective Flight Protection. See deliverables D3.1, D6.1, D5.1 for this issue.

[20] From the game theory point of view, i.e. a strategy that maximises the reward no matter the choices of the other agents. Note that this strategy may or may not be Pareto optimal, i.e. better for everyone on average.

[21] Straightforward only if the airline has a good idea of its costs, which is not always the case.

Moreover, a likely key parameter that will be taken into account is the fairness, or perceived fairness, of the mechanisms. In BEACON we have explored fairness through the different gains, or relative gains, made by the airlines. However, it is not clear yet which indicators should be used exactly in this setting, i.e. which situation should be considered fair. In fact, we expect that the concept of fairness depends strongly on the airline, with airlines with higher costs likely insisting on absolute cost-based indicators being used, whilst airlines with smaller costs may well insist on relative cost being a good indicator of equity. We thus suggest any future research to tackle this problem early on, for instance trying to find a consensus among a large panel of airlines. In fact, we believe that this problem is bigger that the hotspot resolution issue, and that policymakers should clarify the indicator to be used for fairness in general in Europe after consultation with the airlines.

Finally, we provide the table below that summarises the achievements of the project with respect to the objectives set at its beginning. See also the achievements of the project in D1.2.

Objectives	Achievement
"Propose a set of improved flight prioritisation mechanisms that expand current UDPP capabilities."	Three new mechanisms have been proposed, that theoretically allow better cost reduction than UDPP, with incremental difficulty to implement: ISTOP, Credit Mechanism, Auction. Details can be found in D3.1 and D5.1.
"Define new metrics to evaluate the fairness and equity of flight prioritisation mechanisms and validate their appropriateness with AUs."	Three indicators were defined based on previous work from ECTL on fairness and equity. They represent three distinct points of view on equity, presented in D3.1 and D5.2.
"Quantify the impact of 'non-rational' behaviours of AUs on the outcome of the proposed mechanisms, taking advantage of the methods and tools developed in the field of behavioural economics."	Gaming effects were estimated to trigger around 7 percentage points of reduction in economic efficiency and behavioural effects another 5 points of reduction (see D5.2). The models used for behavioural economics were described in D4.1.
"Integrate the insights gained from behavioural economics into an agent-based microsimulation model of the full ECAC network able to capture network effects. The model shall be able to compute a set of key performance indicators (KPIs), including newly developed fairness and equity KPIs, allowing a comprehensive assessment of the new UDPP mechanisms."	Two mechanisms were implemented in the full network simulator. Different agent types were also implemented, as well as KPI capabilities related to Equity and Fairness (see D4.2, D5.1, D5.2).
"Run a set of simulation experiments to evaluate the impact of the new UDPP	We ran extensive experiments in different setups to estimate the efficiency and the equity

EUROPEAN PARTNERSHIP

Co-funded by the European Union

mechanisms on the selected KPIs, taking into account behavioural effects, in order to analyse the advantages and the risks with respect to the current UDPP capabilities."	of the new mechanisms, as well as the gaming and behavioural effects on the performance of the mechanisms, see D4.2, D5.1, D5.2.
"Derive guidelines and methodological recommendations on the further development, validation and deployment of the new UDPP mechanisms that pave the way to a more harmonised and efficient flight prioritisation process across Europe"	We dedicated an entire workpackage to the concept assessment. We finished the project by listing the guidelines for future UDPP mechanisms modelling in D6.2, highlighting the importance of behavioural economics, but also of automation, virtual currencies etc.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

6 References

[1] BEACON Consortium, “D3.1 High-level modelling requirements,” 2021.

[2] BEACON Consortium, “D4.2 Final model results,” 2022.

[3] BEACON Consortium, “D4.1 Behavioural Model Parameter Calibration,” 2021.

[4] BEACON Consortium, “D5.1 First tactical model and results,” 2022.

[5] BEACON Consortium, “D5.2 Final tactical model and results,” 2022.

[6] BEACON Consortium, “D6.1 Intermediate concept assessment report,” 2021.

EUROPEAN PARTNERSHIP

Co-funded by the European Union

7 Acronyms

Acronym	Definition
ATFM	Air traffic flow and capacity management
ATM	Air traffic management
AU	Airspace user
CM	Credit mechanism
FPFS	First-planned first-served
HITL	Human-in-the-loop
HMI	Human-machine interface
ISTOP	Inter-airline slot trading offer provider
KPI	Key performance indicator
LVOC	Low volume operators in constraint
NM	Network Manager
RL	Reinforcement learning
SFP	Selective Flight Protection
SUS	System usability score
UDPP	User Driven Prioritisation Process

EUROPEAN PARTNERSHIP

Co-funded by the European Union