Yevseyeva I, Morisset C, van Moorsel A.

Modeling and analysis of influence power for information security decisions.

***Performance Evaluation* (2016)**

DOI: http://dx.doi.org/10.1016/j.peva.2016.01.003

Modeling and analysis of influence power for information security decisions

Iryna Yevseyeva, Charles Morisset, Aad van Moorsel

PII: S0166-5316(16)00004-3
DOI: http://dx.doi.org/10.1016/j.peva.2016.01.003
Reference: PEVA 1856

To appear in: *Performance Evaluation*

Please cite this article as: I. Yevseyeva, C. Morisset, A. van Moorsel, Modeling and analysis of influence power for information security decisions, *Performance Evaluation* (2016), http://dx.doi.org/10.1016/j.peva.2016.01.003

Modeling and Analysis of Influence Power for Information Security Decisions

Iryna Yevseyeva[1,*], Charles Morisset[1], Aad van Moorsel[1]

[1] *Centre for Cybercrime and Computer Security, School of Computing Science, Newcastle University, Newcastle upon Tyne NE1 7RU, UK*
firstname.lastname@newcastle.ac.uk

Abstract

Users of computing systems and devices frequently make decisions related to information security, e. g., when choosing a password, deciding whether to log into an unfamiliar wireless network, etc. Employers or other stakeholders may have a preference for certain outcomes, without being able to or having a desire to enforce a particular decision. In such situations, systems may build in design nudges to influence the decision making, e. g., by highlighting the employer's preferred solution. In this paper we model influencing in information security to identify which approaches to influencing are most effective and how they can be optimized. To do so, we extend traditional multi-criteria decision analysis models with *modifiable criteria*, to represent the approaches an influencer has available to influence the choice of the decision maker. We also introduce the notion of *influence power*, to characterize the extend to which an influencer can influence decision makers. We illustrate our approach using data from a controlled experiment on techniques to influence which public wireless network users select. This allows us to calculate influence power and identify which design nudges exercise the most influence over user decisions.

Keywords: Information security; security-productivity trade-offs; multicriteria decision analysis; influencing behavior; nudging; influence power.

*Corresponding author

1. Introduction

People continuously make information security decisions: should I use a particular public wireless, should I allow someone's USB to be put in my laptop, how do I choose and memorize passwords? The decisions are often complex, with several objectives to be considered simultaneously, and the optimal decision may very much depend on the specific situation: while using a stranger's USB stick is not advisable, the importance of the job to be completed and/or knowledge about the owner of the USB stick may make it reasonable to use it, despite the associated information security risks.

A simple compliance policy (such as, not to allow USB sticks at all) would be suboptimal. Instead, one would want to allow the owner of the laptop to decide the best course of action. This also emerges in bring your own device (BYOD, [1]), where device owners use their own device for work-related activities. The fact that the user owns the device puts certain restrictions on what the employer can do to implement its preferred security solution. In any case, there are many situations in which the end user is involved in information security decisions that impact other stakeholders.

Although various stakeholders are not in a position to control the outcome of the information security decision, it may be advisable that some stakeholders (e. g., service providers, device vendors, employers) impacted by the end-user decisions, are able to *influence* the decision making, without restricting the freedom of choice of the end-user. Influencing techniques (such as nudging [2]) have been widely used in healthcare and social policies, see e. g. [3], [4], [5], but less so in information security. To establish a sound base to design, evaluate and optimize influencing techniques in information security, we need a formal and coherent framework to analyze and evaluate influencing.

In our earlier work [6], we identified an agent-based model that allows one to reason conceptually about influencing in information security. We introduced the general notion of optimal influencing policy, taking into account uncertainty of the environment and the fact that agents have partial and differing ability to observe that environment. The paper shows that end-user decisions under influencing may outperform the decisions made by either the end-user or influencer alone (in terms of [6], it shows that 'soft enforcement' can be better than weak or strong enforcement).

In this paper we are after a more operational model, that allows us to identify for real-life scenarios which approaches to influencing are most ef-

fective and how these approaches can be optimized. To this end, we apply and extend well-known models from multi-criteria decision analysis, a well-understood and frequently used approach to modeling human decision making, see e. g. [7]. In our model, both decision-maker and influencer make decisions governed by multi-attribute value theory [8]. To represent influencing we introduce *modifiable criteria*. Modifiable criteria reflect the impact an influencer has on the decision maker but do not change the available alternatives to chose from. For instance, if the influencer uses colors when presenting alternatives, the coloring does not change which options are available but does influence the value the decision makers associates with its decision criteria.

We also introduce the notion of *influence power*. Influence power expresses the extend to which a user is susceptible to being influenced, in terms of individual criteria. Vice versa, by adding a cost function to modifying criteria, influence power also allows one to express the effort needed by the influencer to successfully change the user's decision. That is, influence power allows one to evaluate and compare the effort needed in influencing approaches and therefore provides a tool to optimize the design and application of an influencing approach.

To illustrate the use of the concepts introduced in this paper we parameterize a model using data from a controlled experiment in WiFi network selection [9]. The experiment provides data about a number of design nudges which aimed at influencing which WiFi network would be selected–these design nudges include coloring of available networks, changing the order they are presented, etc. We show how the influence power of a particular criterion can be computed and how much influence power is needed to change the choices of participants with different preferences. We are able to determine whether the influencer can change the choice of decision makers by changing only one criterion at a time or a (sub)set of modifiable criteria. Such insights can guide the designer of nudges and improve the effectiveness of techniques for influencing choice.

The work presented in this paper advances most elements of our earlier work in [10], which was the starting point for this special issue paper. The formalization is extended with the notion of influence power and with that of modifiable criteria. The analysis using experimental data from a study of nudging in public Wi-Fi network selection has been extended considerably, particularly through results in Section 5. We also extended the related work discussion, especially targeting the community of probabilistic system mod-

elers, and we provide a much deeper discussion of remaining challenges and opportunities.

The paper is organized as follows. Before introducing our modeling approach, Section 2 discusses key elements of the state of the art in influencing techniques, drawing from literature in various disciplines. It also discusses related modeling approaches (particularly Markov decision models and reinforcement learning). Section 3 introduces our structured approach to modeling influencing, including the concepts of influence power and modifiable criteria. Section 5 provides the data analysis for the WiFi network selection experiment, preceded by an explanation of the case study specifics in Section 4. We discuss gained insights in Section 6, referring to a number of interesting issues for further research and refinement, including intuitive decision making strategies, influencing through a 'decoy' alternative, and influencing a larger population of users. Section 6 provides the final conclusions.

2. Background and Related Work

Before presenting our approach to modeling influence, Section 2.1 provides a background discussion of influencing techniques in general and Section 2.2 positions our approach of using multi-criteria decision making in the context of Markov decision modeling and reinforcement learning.

2.1. Influencing Decision Making

Influencing decision making has attracted attention of researchers from many fields, including psychology, behavioral economics, marketing, health and, more recently, privacy and security. The research in these varying domains all discuss, from their own disciplinary perspective, the factors that influence choices and the design of interventions to try to influence decisions. Examples of such interventions are wide-spread, for instance interventions that aim to change smoking or eating habits [4]. In marketing, influencing is part and parcel of anything it aims to achieve, whether it is to improve product sales or protect the customer [11].

In recent times, policy makers have become increasingly interested in a specific instance of influencing, namely *nudging*, as made popular by the work of Thaler and Sunstein [2]. Nudging refers to the design of a *choice architecture*, the manner in which choices are presented or framed. The aim of the choice architecture is to influence the decision maker (in the desired direction), without restricting the freedom of choice. Even small changes in

the presentation may influence final choices of decision makers, as shown by [12], [13]. A wide variety of practical examples of influencing health and social decisions exists [2], [3], [4], [5]. For instance, it was shown that rearranging menu items in a cafeteria may increase consumption of a particular (healthy) item by up to 25% [2].

Recently, researchers have started to study nudging in the context of security and privacy decision making [9] [14], [15], [16]; the case study in this paper is an example of nudging from [9]. In security, nudging is potentially a soft and flexible alternative to the more common approach of strict compliance to a security policy. In the introduction we already indicated that nudging may be appropriate in a variety of information security scenarios, either because the end user has the right to make the decision, or is best positioned to make an informed decision. At the same time, nudging in information security may be less straightforward than that in health or other domains. One of the challenges in information security is that the optimal decision may be a trade-off between various concerns, including ease of use, performance, risk appetite and security. That is, it is not always obvious in which direction to nudge. However, as shown in [15] for WiFi selection and in [14], [16] for privacy, if an objective is agreed, then influencing can be an effective approach to improving security and privacy without restricting the user's choices.

In [15] we introduced the SCENE design process for choice architecture, based on the MINDSPACE framework [17], which targets approaches influencing beyond nudging alone. MINDSPACE identifies a number of categories of phenomena that explain influencing, including messenger, norms, salience, priming, affect, commitment and ego. (For example, salience highlights importance of particular choices, and priming frames solutions in either positive or negative context.) The design process in [15] then provides a process for companies to explore the creation of nudges using the MINDSPACE framework, as part of creative workshops. [15] also explains the role of models in various stages of the design, in particular in optimizing and evaluating the choice architecture. This paper provides some of the modeling techniques that can be used as part of the SCENE design process.

2.2. Quantitative Modeling of Information Security Decision Making

In recent years, several approaches to modeling human information security decision making have been proposed [18] [19]. These belong to the

category of discrete-event dynamics systems, in which the state of a system changes over time, at discrete points in time. Models belonging to that class include Markov chains and other variants of Markov processes, as well as models expressed through stochastic process algebras, e. g., [20]. With respect to decision making, two analysis approaches are of particular relevance, namely Markov decision processes and reinforcement learning, which we both discuss briefly.

A *Markov decision process (MDP)* [21] represents a finite set of states with transitions from one state to another through some actions/decisions that are taken over time. A transition from state to state given a decision is probabilistic. Probabilities of all transitions to all reachable states are assumed to be known; also for each of such transition a reward is associated. Then, a policy is a function taken over all decisions from the initial to some final state. The goal of optimizing MDP is to find a sequence of decisions to take such that it maximizes the total reward. Each state of an MDP depends on the previous state only.

Reinforcement learning [22] models operate in a potentially uncertain environment with a sequence of actions/decisions to be taken over time based on experience obtained in earlier steps. Rewards of decision sequences may be only partially known or unknown at all. Behavior of multiple agents can be modeled assuming that agents observe the environment. Each decision of an agent depends on the previous decisions taken by the agent or other agents and the current observation of environment, which may be full or partial, including information on rewards associated with each potential decision to be taken. The process of decision making in reinforcement learning can be modeled with an MDP, although large state spaces can be expected as a result of the fact that actions depend on the history of decisions. In this case, simulation approaches must be used to deal with the state space explosion

The modeling approach in this paper differs fundamentally from the above approaches because it does not model the behavior and dynamics of the system itself. Instead, it models the factors that influence a decision, effectively at some non-specified moment in time (there is no explicit notion of time in these models). However, one can imagine that merging approaches may be useful in future expansion of the ideas proposed in this paper. For instance, when connecting to a public Wi-Fi, see Section 4, the above approaches would allow one to model the environment, typical user behavior, as well as sequences of preceding decisions. For instance, knowing that the decision maker is using Virtual Private Network, which allows communication

through a public Wi-Fi with all benefits of a private one, including security, would have implications for the best implementation of a nudge, since connecting to an unsecured Wi-Fi will be less harmful.

3. Influencing in Decision Making Models for Multiple Criteria

Multi-Criteria Decision Making refers to a collection of operations research techniques used to formally model and reason about human decision-making in situations in which multiple and possibly conflicting criteria need to be considered when choosing an alternative. Multi-criteria decision making is used to search for those alternatives that represent the best trade-off between criteria. Commonly, two branches of multi-criteria decision making are identified in the literature [23], namely multi-objective optimization and multi-criteria decision analysis, which are differentiated by how alternatives are treated. In the former case, the alternatives from which one can chose are not available in advance and are discovered in the process of optimization within the space of decision variables (infinite, continuous or large, combinatorial discrete) and restricted by constraints. In the later case, the alternatives set is available in advance and the space of alternatives is usually discrete, feasible, countable and small enough to be compared pairwise by the decision maker [23].

When considering multiple criteria when choosing an alternative, one can identify the candidate set of optimal alternatives, called the Pareto set [24], [25]. Alternatives in the Pareto set are non-dominated solutions, which means that for each alternative in the set no other alternative exists that is better with respect to all criteria. The set of Pareto alternatives can be large and even combinatorial in the size of the problem, and if that is the case, approaches are needed to simplify the selection of the optimal alternative within the set. One approach to determine the optimum is provided by using utility functions or one of its variants, which associate a single quantity with each alternative in the Pareto set. In this paper we will use one such approach, namely Multi-Attribute Value Theory (MAVT, [8]).

3.1. Multi-Attribute Value Theory: Basic Model

MAVT resolves dealing with trade-offs by introducing a value function, which takes values associated with criteria, and returns a single value for each alternative. Then an alternative is considered to be better than the other one if its value is better. MAVT assumes that decision makers are able

to assess performance of alternatives on individual criteria and to specify relative importance of criteria (or criteria weights). Then, the value function determines which alternative is best, e.g. by computing a weighted sum of individual criteria values. Value functions are specific to the particular decision maker and reflect decision maker's preferences. MAVT's value function makes it a 'compensatory technique', allowing the construction of functions that smaller values on a subset of criteria to balance with some large values on others. Mathematically, MAVT is closely related to Multi-Attribute Utility Theory [7], which has a probabilistic interpretation and is based on expected utility theory with some strong technical assumptions related to comparability, transitivity, continuity, and independence of outcomes (that assumes independence of criteria). Both multi-attribute value and utility theory are attractive because of their sound theoretical foundation.

We now provide a formalization of MAVT so that we can introduce constructs that help us analyze influencing. A decision maker needs to select an alternative a from a set $\mathcal{A}$. Alternatives are evaluated by decision makers using criteria from the set of all criteria $\mathcal{G}$. Each criterion $g \in \mathcal{G}$ comes with an ordinal or ordered categorical scale $\mathcal{K}_g$ that defines the possible values the criterion can take. Typical scales include real numbers, intervals, ratios, binary or qualitative indicators (such as 'High', 'Green', etc.). Qualitative scales are often mapped into real numbers $\mathcal{K}_g \rightarrow \mathbb{R}$ to allow aggregation, and we will indeed assume that all value are real values in what follows. For later use, the minimal and maximal values of on the scale for g are denoted as $g^{\min}$ and $g^{\max}$, respectively.

We illustrate above for the case study discussed in Section 4. Assume one needs to select a public Wi-Fi to perform some task and two networks are available: $\mathcal{A} = \{s, f\}$, where s is a secured network with a weak signal and f a non-secured network with a strong signal. Whether the network is secured and signal strength can each be represented by a criterion: $\mathcal{G} = \{t, r\}$, with t representing trust (i. e., secured network) and r strength. We may decide to use for both of them the same scale $\mathcal{K}_t = \mathcal{K}_r = \{0, 1, 2\}$ (with higher values preferred).

Values associated with criteria differ across alternatives, so we say that values are defined by a *partial value function* $v_g : \mathcal{A} \rightarrow \mathcal{K}_g$, which is also referred in literature as a marginal value function. We write V_g for all possible v_g functions, and $V_{\mathcal{G}} = \prod_{g \in \mathcal{G}} V_g$ for the cartesian product of all partial value functions. When no confusion can arise, we write $v(a) = (v_{g_1}, \ldots, v_{g_m})$ for the vector of partial value functions belonging with the criteria of alternative $a \in$

$\mathcal{A}$. To facilitate aggregation in a single value, criteria values are normalized, to provide a fair basis for comparison. We write the normalization function as $n_g : \mathcal{K}_g \to [0, 1]$, for $g \in \mathcal{G}$.

Preferences are encoded using criteria weights, which express relative importance and trade-offs between criteria, essentially defining how many units of one criterion can be traded-off for a unit of another criterion. Criteria weights are defined through a vector $w(a)$, with each criterion weight $w_g : \mathcal{A} \to [0, 1]$ such that $\sum_{w_g(a) \in w} w_g(a) = 1$, for all $a \in \mathcal{A}$. Although not strictly necessary in MAVT, we use the convention for weights to be between 0 and 1 as in utility theory (where weights are interpreted as probabilities).

We are now in a position to provide a formal characterization for MAVT:

Multi-Attribute Value Theory Model. A MAVT model is a tuple $M = (\mathcal{A}, \mathcal{G}, \mathcal{K}, V_{\mathcal{G}}, n, W)$, where $\mathcal{A}$ is a set of alternatives, $\mathcal{G}$ is a set of criteria, $\mathcal{K}$ is a set of criteria scales, $V_{\mathcal{G}}$ is a set of partial value functions with $v_g : \mathcal{A} \to \mathcal{K}_g$ for each pair $\{a, g\}$, n is a set of normalization functions with $n_g : \mathcal{K}_g \to [0, 1]$ for each g, and W is a set of vectors of weights $w(a)$, for each $a \in \mathcal{A}$, with $w_g(a) \in [0, 1]$ and such that for each $a \in \mathcal{A}$, $\sum_{w_g(a) \in w(a)} w_g(a) = 1$.

To aggregate normalized criteria values for each alternative various candidate aggregation function can be thought of, e. g. multiplicative, additive or some combination thereof. In this paper we will use the most straightforward aggregating value function, which fits naturally with our use of weights, namely the weighted sum [7]:

Value Function. Given a model $M = (\mathcal{A}, \mathcal{G}, \mathcal{K}, V_{\mathcal{G}}, n, W)$, the value function v of an alternative $a \in \mathcal{A}$, is defined as:

$$\mathsf{v}(a, w, v) = \sum_{g \in \mathcal{G}} w_g(a) \cdot n_g(v_g(a)). \tag{1}$$

The preferred alternative then is the alternative $a \in \mathcal{A}$ that has maximum valued value function. We note that in the above equation including w and v in $\mathsf{v}(a, w, v)$ is redundant, since a alone specifies the value and weight functions already. However, we will need to include w and v as parameters in $\mathsf{v}(a, w, v)$ to facilitate the discussion on (optimal) modifiable criteria later on.

3.2. MAVT for Influencing Information Security Decisions: Optimal Policy and Impact

When analysis influencing, we need to consider two different parties, the decision maker and the influencer–a typical example of the two parties would be the user of a device as decision maker, and the employer as the influencer. There is a MAVT model associated with each of the parties, one for the decision maker and one with the influencer. We use the subscript DM for the decision maker and I for the influencer.

One particular case is where the decision maker and influencer exhibit the same preferences i. e., the same weights $w_I = w_{DM}$ and the same set of partial value functions: $v_I = v_{DM}$. Since the decision maker already choses the same alternative as the influencer, influencing would not be necessary, and could even be counter productive. Reality would typically be different, with partial value functions $v_I \neq v_{DM}$ and weight vectors $w_I \neq w_{DM}$ different for the two parties.

We now want to map MAVT terminology and concepts on those derived for influencing in [10] and [26]. First, we introduce the notion of *policy* $\pi(w, v) \in \mathcal{A}$, intuitively referring to something similar as an information security policy, that is, the information security rules one would want to follow or impose. Given a vector of partial value functions v, and a weight vector w, the policy $\pi(w, v) \in \mathcal{A}$ of a decision maker is defined as [10]:

$$\pi(w, v) = \arg\max_{a \in \mathcal{A}} \quad \mathsf{v}(a, w, v). \tag{2}$$

It will become apparent in the following subsection that this notation for impact is useful when reasoning about modifiable criteria modification. Note that in order for a decision maker to be deterministic, we assume the existence of an arbitrary ordering over alternatives, so that if there are several alternatives maximizing the value function, the decision maker selects the highest one according to that ordering.

The *impact function* ρ was introduced in [26] to express the benefit of making certain decisions, and again we introduce notation to express in the next section impact of influencing attempts. In general the impact function can be defined in many different ways (for instance, through an access control policy stating which alternatives are secure [26]), but in this paper we have already defined a version of impact through the value function v of M. However, in the context of information security, we want to clearly distinguish

the alternatives, so that there are 'good' and 'bad' alternatives. Hence, for an alternative a, we define the impact function as:

$$\rho(a, w, v) = \begin{cases} 1 & \text{if } \mathsf{v}(a, w, v) \geq \mathsf{v}(a', w, v) \text{ for all } a' \in \mathcal{A}, \\ 0 & \text{otherwise.} \end{cases}$$

In other words, an alternative has an impact if, and only if, it is maximal according to the value function. Note that more complex impact functions can be considered, for instance, when different levels of value can be defined.

3.3. Introducing Influence in MAVT

This subsection extends MAVT to represent and analyze influencing. Our approach is as follows. Given the set of alternatives $\mathcal{A}$ of the decision maker, attempts to influence by the influencer inflict a change in some of the partial value functions of criteria associated with alternative $a \in \mathcal{A}$. However, influencing does not alter the set $\mathcal{A}$, still the same alternatives are available. We say, given the set of criteria $\mathcal{G}$, a subset of *modifiable criteria* $\mathcal{M} \subseteq \mathcal{G}$ is available to the influencer. The exact subset depends of course on the problem at hand, but intuitively, it corresponds anything the decision maker takes into account that is under the (partial) control of the influencer.

We say that there is *influence*, if and only if, the decision maker decides differently than when not being influenced, that is, when the optimal policies before and after modification are such that $\pi(w, v) \neq \pi(w, v')$. Here, we use v' to express modifications to the partial value functions of criteria (from v to v').

As an example, we return to the WiFi network selection example. Earlier in this section we introduced two alternatives $\mathcal{A} = \{f, s\}$, and two criteria $\mathcal{G} = \{t, r\}$, corresponding to network trust and strengths, respectively. The influencer now may wish to attempt to influence the choice of the decision maker by modify the displayed number of bars available for strength, e. g., from $\{0, 1, 2\}$ to $\{0, 1, 2, 3, 4\}$. Or the influencer may change the color of the displayed name of the network, again to influence selection. In both cases, the decision maker sees a different presentation of the alternatives, which then is reflected in a change in the value the decision maker associates with a criterion.

To determine which modification is the best for the influencer we first define the set of all modifications of criterion values available to the influencer, and then take from that set the modification with the highest impact. Given

a vector of partial value functions v, the set of all possible modification functions is:

$$P_{\mathcal{M}}(v) = \{v' \mid \forall g \in (\mathcal{G} \setminus \mathcal{M})\ v'[g] = v[g]\}.$$

In the case study, we will be able to enumerate all possible modification that are possible. One can also chose to define further restrictions on $P_{\mathcal{M}}$, for instance, reflecting a limit on the change in the values of criteria (e. g., the value of a criterion can only be incremented or decremented by a given factor). The optimal modification possible by an influencer over a decision maker can now be defined as follows:

Definition 1 (Optimal Modifications)**.** *Given an influencer with a weights function w_I and a vector of partial value functions v_I, and a decision maker with a vector of weights w_{DM} and a vector of partial value functions v_{DM}, the vectors of optimal modified partial value functions for the decision maker are given by:*

$$\mathsf{opt}(w_I, v_I, w_{DM}, v_{DM}) = \arg \max_{v'_{DM} \in P_{\mathcal{M}}(v_{DM})} \rho(\pi(w_{DM}, v'_{DM}), w_I, v_I). \quad (3)$$

In words, this says that the optimal modification v'_{DM} is the one that results in a reaction of the decision maker that is of the most benefit to the influencer.

There may be several vectors of optimal modifications that result in same optimal policy. To select one of them we introduce cost $c(v_{DM}, v'_{DM})$, which is associated with the amount of effort needed by the influencer to achieve the optimal change. The influencer will be interested in finding a vector of optimal modifications that require minimal effort, which we call here *Vector of Optimal Modifications*.

Definition 2 (Vector of Optimal Modifications)**.** *Given the same model elements as in the previous definition, the Vector of Optimal Modifications is defined as:*

$$\mathsf{opt}^*(w_I, v_I, w_{DM}, v_{DM}) = \arg \min_{v'_{DM} \in \mathsf{opt}(w_I, v_I, w_{DM}, v_{DM})} c(v_{DM}, v'_{DM}). \quad (4)$$

Associating cost/effort with modifications provides us with a way of quantifying the effectiveness of different approaches to influencing. This leads to the introduction of the second key novel concept (after modifiable criteria), namely Influence Power. Influence Power is defined through the cost of performing optimal modifications computed with (3):

Definition 3 (Influence Power)**.** *Given the same model elements as in the previous definitions, the* Influence Power IP *is defined as:*

$$\mathsf{IP}(w_I, v_I, w_{DM}, v_{DM}) = \min_{v'_{DM} \in \mathsf{opt}^*(w_I, v_I, w_{DM}, v_{DM})} c(v_{DM}, v'_{DM}). \tag{5}$$

3.4. Influence Power for the Case Study

To close this section we introduced variants of the notion of Influence Power useful for the discussion of the case study in Section 4. In the case study, all scales $\mathcal{K}$ are discrete, and as we already remarked, this allows enumeration of the elements of the set $P_{\mathcal{M}}(v)$ of all possible modification. It also provides a natural way to assign cost to modifications, namely by counting the number of changes needed to go from v to v'. It is then more illustrative to look at influence power relative to the the maximum number of changes possible. For the case study, the scales all have the same direction (higher is better), which allows us to define *relative influence power* as follows:

Definition 4 (Relative Influence Power)**.** *Given the same model elements as in the previous definitions, and with v_{DM}^{min} and v_{DM}^{max} vectors of partial value functions with all the smallest and all the largest values on all criteria scales, respectively. Then,* Relative Influence Power *is defined as:*

$$\mathsf{RIP}(w_I, v_I, w_{DM}, v_{DM}) = \frac{\mathsf{IP}(w_I, v_I, w_{DM}, v_{DM})}{c(v_{DM}^{min}, v_{DM}^{max})} \cdot 100\%. \tag{6}$$

Note that in the above setting (since it intuitively represents 'effort needed' to change the decision maker's choice) a low value of RIP is better. We also note that it is very well possible that the optimal modification for the influencer is to not make any changes, i. e., $v_{DM} = v'_{DM}$. Formally, the influencer should not make use of any of the influencing techniques if one of the two following statements holds:

$$\text{if } \rho(\pi(w_{DM}, v_{DM}), w_I, v_I) = 1, \text{ then } v_{DM} \in \mathsf{opt}(w_I, v_I, w_{DM}, v_{DM}),$$

or:

$$\text{if } \nexists v'_{DM} \text{ s.t. } \rho(\pi(w_{DM}, v'_{DM}), w_I, v_I) = 1, \text{ then } v_{DM} \in \mathsf{opt}(w_I, v_I, w_{DM}, v_{DM}).$$

4. Case Study: Selection of a Wi-Fi Network

In this section we illustrate the modeling ideas presented in this paper. We present the core of the decision model we constructed when a user has to decide which Wi-Fi network to select in a public space. In this section we do not yet use the data from the experiment, that is done in Section 5. The MAVT model we use in this section provides the core of the full model needed in Section 5.

4.1. Selection of a Wi-Fi Network with MAVT

The decision maker chooses between two different public wireless networks $\mathcal{A} = \{s, f\}$: s is a secure Wi-Fi with weak signal; f is a Wi-Fi with strong signal, but not necessarily secure. This illustrates the trade-off between security and productivity/usability. The set of criteria is $\mathcal{G} = \{t, r, l\}$, indicating the *trust* or security t of the network, its *strength* r and the *color* l in which the network name is displayed. The scales for the trust and strength criteria are defined as $\mathcal{K}_t = \mathcal{K}_r = \{0, 1, 2\}$ (higher is better). For the color criterion, the scale is $\mathcal{K}_l = \{R, N, G\}$, corresponding to red, neutral and green font colors.

In the language of the Wi-Fi application, trust t relates to the level of security, e. g. whether the Wired Equivalent Privacy (WEP) protocol is used and/or whether it is in the whitelist of the networks recognized by the decision maker's device. More sophisticated evaluation of trust may take into account other aspects, e. g., current location of an employee [27]. Strength r of a network can be associated with the number of bars showing the strength of the signal, and one may expect that people would chose networks with high signal strength. Color l indicates the font color in which the network name is displayed, through a traffic light system [28]. This associates colors to emotions: red color – to danger, amber – to attention, and green – to no danger, and is a common ingredient in design nudges. For instance, in [14] the traffic light approach is used to nudge individuals away from privacy-invasive applications in.

The decision maker defines the criteria weights, e. g. $w = (0.5; 0.3; 0.2)$, or, alternatively, weights can be derived from observed decision maker choices. Weights $w = (0.5; 0.3; 0.2)$ can for instance be interpreted as follows: connecting to a trusted Wi-Fi is more important for the decision maker than choosing a Wi-Fi with strong signal. The color in which a Wi-Fi name is displayed, is less significant for the decision maker than the two other criteria,

Table 1: Decision matrix corresponding to $v = [s \mapsto (1, 1, N), f \mapsto (0, 2, N)]$.

	Criteria $\mathcal{G}$		
	Trust (t)	Strength (r)	Color (l)
Weight Vector w	0.5	0.3	0.2
Alternatives $\mathcal{A}$			
Safe (s)	1	1	N
Fast (f)	0	2	N

etc.

In the following, for the sake of compactness, it is written $v = [s \mapsto (v_1, v_2, v_3), f \mapsto (v_4, v_5, v_6)]$, associating network s with a trust of v_1, a strength of v_2 and a color of v_3 values, respectively; and network f with a trust of v_4, a strength of v_5 and a color of v_6 values, respectively. Table 1 represents the traditional decision matrix [7], corresponding to the compact notation $v = [s \mapsto (v_1, v_2, v_3), f \mapsto (v_4, v_5, v_6)]$.

In Table 1, available alternatives s and f (rows in the lower part) with their values on criteria t, r, and l (columns) are presented. In addition, criteria weights are given in the headings of columns. We assume a linear normalization function of the following form:

$$n_g(v_g(a)) = \frac{v_g(a) - g^{min}}{g^{max} - g^{min}}. \tag{7}$$

E.g. for scales $\mathcal{K}_t = \{0, 1, 2\}$, $\mathcal{K}_r = \{0, 1, 2\}$, $\mathcal{K}_l = \{R, N, G\}$, normalized scales all equal $\mathcal{K}_t^n = \mathcal{K}_r^n = \mathcal{K}_l^n = \{0, 0.5, 1\}$. So, we now can compute the value functions (assumed to be additive sum as in Eq. (1)) for alternatives s and f:

$$\mathsf{v}(s, w, v) = 0.5 * 0.5 + 0.3 * 0.5 + 0.2 * 0.5 = 0.5$$
$$\mathsf{v}(f, w, v) = 0.5 * 0 + 0.3 * 1 + 0.2 * 0.5 = 0.4.$$

In this case, in terms of Eq. (2), the decision maker will use policy $\pi(w, v) = s$, that is, the secure network will be chosen.

To illustrate decision evaluation scenarios for the case of public Wi-Fi selection, let us consider the influencer and the decision maker having the same weight vectors $w_I = w_{DM} = (0.5; 0.3; 0.2)$. However, they have different

partial value functions: $v_I = [s \mapsto (1, 1, N), f \mapsto (0, 2, N)]$ for the influencer and $v_{DM} = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$ for the decision maker. Indeed, the decision maker considers the alternative f as being more trusted, with $v_{DM}[t](f) = 1$, when compared to the influencer, which assigns to it a smaller trust value with $v_I[t](f) = 0$. This difference results in different utilities of the alternatives $\mathsf{v}(s, w_{DM}, v_{DM}) = 0.5$ and $\mathsf{v}(f, w_{DM}, v_{DM}) = 0.65$, and leads to the decision maker choosing $\pi(w_{DM}, v_{DM}) = f$. However, $\rho(f, w_I, v_I) = 0$, meaning that the decision maker selects an alternative that is suboptimal for the influencer.

4.2. Influencing Selection of a Wi-Fi Network

Let us consider as subset of modifiable criteria $\mathcal{M} = \{l\}$, i. e., only the color in which network's name is displayed can be modified. Assuming an influencer has a set of criteria weights $w_I = (0.5; 0.4; 0.1)$ and a set of partial value functions $v_I = [s \mapsto (1, 1, N), f \mapsto (0, 2, N)]$, then the impact of alternatives s and f can be computed for the influencer as follows, $\rho(s, w_I, v_I) = 1$ and $\rho(f, w_I, v_I) = 0$, respectively. In other words, the influencer wants to influence the decision maker towards selecting a more secure Wi-Fi s. The decision maker has criteria weights $w_{DM} = (0.3; 0.5; 0.2)$ and partial values functions $v_{DM} = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$. This leads to the decision maker choosing a fast network $\pi(w_{DM}, v_{DM}) = f$.

Since $\mathcal{M} = \{l\}$, only the color criterion value can be modified. Table 2 details all the possible cases of modifying color, where we write v_{DM}^{xy} for the partial value function $v_{DM}^{xy} = [s \mapsto (1, 1, x), f \mapsto (1, 2, y)]$ of decision maker when modification to the color was applied. To break ties, we assume that when s and f have the same value for their value function, the decision maker selects f by default.

The vector of optimal modifications is $\mathsf{opt}(w_I, v_I, w_{DM}, v_{DM}) = [s \mapsto (1, 1, G), f \mapsto (1, 2, R)]$, i. e., changing the display color of network s to green, and that of f to red, results in the decision maker to change their choice from the alternative f to the alternative s. The decision maker then selects the alternative preferred by the influencer. Note that here there is only one vector of optimal modifications.

The impact of a modification depends on the set of non-modifiable criteria $\{t, r\}$. See Table 3, for the same set of partial value functions $v_{DM} = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$ but different set of weights $w_{DM} = (0; 0.8; 0.2)$. Table 3 demonstrates that if a decision maker does not care about the trust of the

Table 2: Impact of all possible modifications to the color criterion, $v_{DM} = [s \mapsto (1,1,x), f \mapsto (1,2,y)]$, $w_{DM} = (0.3; 0.5; 0.2)$. The decision maker only selects network s if the network s is displayed in green ($x = G$) and the network f is displayed in red ($y = R$), as highlighted in bold.

v_{DM}^{xy}	$\mathsf{v}_{DM}(s, w_{DM}, v_{DM}^{xy})$	$\mathsf{v}_{DM}(f, w_{DM}, v_{DM}^{xy})$	$\pi(w_{DM}, v_{DM}^{xy}) = a$	$\rho(a, w_{DM}, v_{DM}^{xy})$
v_{DM}^{NN}	0.5	0.6	f	0
v_{DM}^{NR}	0.5	0.5	f	0
v_{DM}^{NG}	0.5	0.7	f	0
v_{DM}^{GN}	0.6	0.6	f	0
v_{DM}^{RN}	0.4	0.6	f	0
$\mathbf{v_{DM}^{GR}}$	**0.6**	**0.5**	**s**	**1**
v_{DM}^{RG}	0.4	0.7	f	0
v_{DM}^{RR}	0.4	0.5	f	0
v_{DM}^{GG}	0.6	0.7	f	0

network, there are no modifiable criteria that are effective when trying to make the decision maker select the more secure network.

The relative influence power RIP, expressing the cost of optimal modifications as a fraction of maximal influence power value, can be computed as follows. Criterion t can take three values (0, 1, or 2), thus allowing for four changes, two in each direction: from 0 to 1, from 1 to 2, from 2 to 1 and from 1 to 0. Criteria r and l also allow for 4 changes, thus resulting in 12 for the total amount of possible changes. Assume the colors initially equal to N for both networks, then changing the choice of the decision maker requires 2 grades changes (from neutral to green and from neutral to red, respectively) $\mathsf{RIP} = \frac{2}{12} = 16,66\%$.

5. Experimental Results for Influencing Wi-Fi Network Selection

We apply our modeling approach to the problem of WiFi network selection in experimental setting and study whether we can influence which network users will choose when they are in public spaces. Our analysis uses data from the controlled experiment reported in [9] and further analyzed in [29] and [30], in which a number of design nudges were introduced in order to influence the user's network selection. The set-up is described in Section 5.1. We then explain how we formulated the Wi-Fi network selection problem in

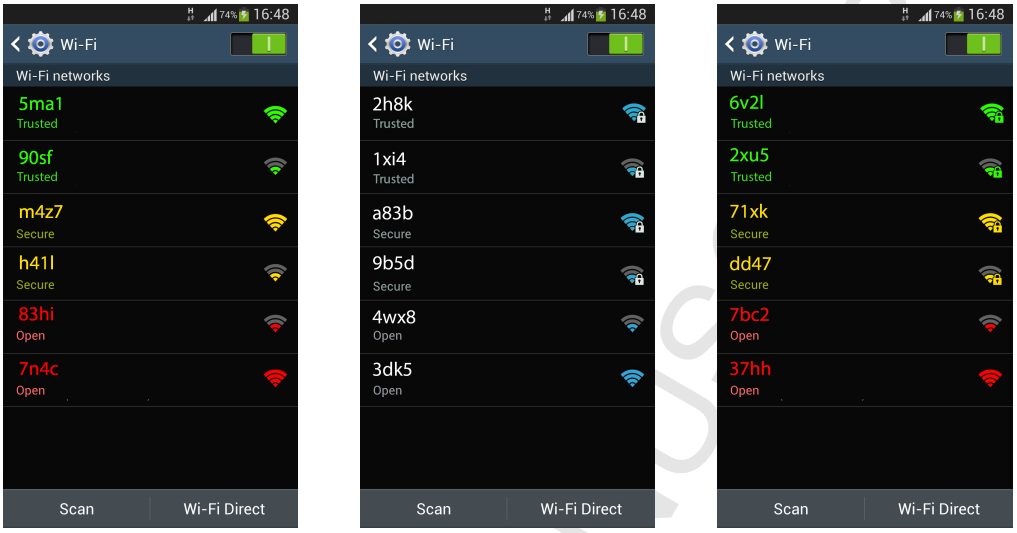
Table 3: Impact of all possible modifications to the color criterion, $v_{DM} = [s \mapsto (1,1,x), f \mapsto (1,2,y)]$, $w_{DM} = (0.0; 0.8; 0.2)$. Irrespective of the influencing attempt (through coloring the displayed network name), the decision maker will not select secure network s since it associates zero weight to criterion t (trust).

v_{DM}^{xy}	$\mathsf{v}_{DM}(s, w_{DM}, v_{DM}^{xy})$	$\mathsf{v}_{DM}(f, w_{DM}, v_{DM}^{xy})$	$\pi(w_{DM}, v_{DM}^{xy}) = a$	$\rho(a, w_{DM}, v_{DM}^{xy})$
v_{DM}^{NN}	0.5	0.9	f	0
v_{DM}^{NR}	0.5	0.8	f	0
v_{DM}^{NG}	0.5	1	f	0
v_{DM}^{GN}	0.6	0.9	f	0
v_{DM}^{RN}	0.4	0.9	f	0
v_{DM}^{GR}	0.6	0.8	f	0
v_{DM}^{RG}	0.4	1	f	0
v_{DM}^{RR}	0.4	0.8	f	0
v_{DM}^{GG}	0.6	1	f	0

MAVT in Section 5.2. Using the notion of modifiable criteria, we identify in Section 5.3 which criteria are most suitable for successful influencing. Assuming each individual criterion as modifiable we compute influence power needed for changing choices. From this investigation we concluded that it is not always possible to influence decision makers with subset of modifiable criteria and possibilities of influencer are limited. In Section 5.4 we further explore how much relative influence power is needed for various groups of decision makers to make them changing their choices. From this study we concluded that the amount of influence power needed for various screenshots depends on the number of criteria used at that screenshot.

5.1. Design of Experiment

The participants were recruited at the Newcastle University. Altogether 34 individuals participated, among which were computing, non-computing students, post-doctoral researchers and several professionals. The participants were asked to imagine a need for a document to be submitted urgently using a publicly available wireless network in a cafe. A set of 6 mobile phone screenshots were presented to participants in random order with 6 Wi-Fi networks displayed in each screenshot. The screenshots are given in Figure 1 and Figure 2. Participants were asked to rank networks according to the order in which they would try to connect to the networks, assuming they



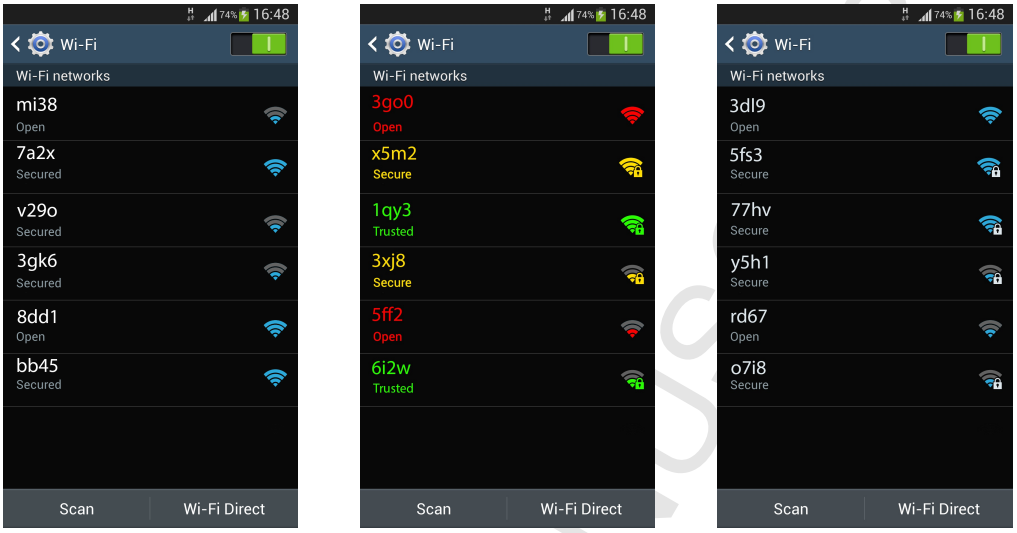
(a) OCn Screenshot (b) OWp Screenshot (c) OCp Screenshot

Figure 1: Screenshots with Wi-Fi's ordered according to their utility to influencer [9]

have the password to access every Wi-Fi available. The network names were composed from randomly generated alphabetic characters and numbers to avoid bias due to perceived familiarity of network names (see [27] for a study of this effect).

The names of screenshots encode the manner of presentation, which relates closely to the criteria to be used in the MAVT model of Section 5.2. For the six names of the screenshots, the first letter stands for 'Order' ('O' for order being used and 'R' for not used); the second letter stands for 'Color' ('C' for a traffic light color scheme being used and 'W' for it not being used, in which case all Wi-Fi names are displayed in white); and the third letter stands for 'Padlock', which indicates usage of the padlock symbol ('p' for being used, 'n' for not being used).

Table 4 provides the named screenshots and also indicates which criteria can be studied through the respective screenshots. 'Order' indicates that Wi-Fi's were initially pre-ordered, with the most secure networks located at the top of the list, see screenshots OCn, OWp, OCp in Figure 1. Otherwise, Wi-Fi's are presented in random order, see screenshots RCp, RWp, RWn in Figure 2. 'Color' refers to the display of Wi-Fi names in different colors: green, neutral, amber or red. We also have text under the names



(a) RWn Screenshot (b) RCp Screenshot (c) RWp Screenshot

Figure 2: Screenshot with randomly presented Wi-Fi's [9]

of networks, either 'Trusted', 'Secure' or 'Open'. Both 'Color' and 'Trust' are used in OCn, OCp, RCp screenshots and not used in OWn, RWn, RWp screenshots. 'Padlock' refers to the padlock graphical symbol, and is used in some screenshots: OCp, OWp, OCp, while no padlock symbol is used in other screenshots OCn, OWn, RWn. 'Strength' is indicated by the number of solid bars of the usual signal graphic, reflecting the strength of the Wi-Fi signal, and is present in all screenshots.

The experiment has been used in a number of papers to discuss the psychological and information security implications. In particular, in [9] the application for nudging users was presented for android phones. In the same work it was investigated to what extend different features of Wi-Fi's, such as color, place in the ordered list and presence of the padlock symbol, may nudge decision makers towards selecting more or less secured networks. In addition, follow up studies explored how individual differences, such as technical self-efficacy, perceived controllability and vulnerability to risk, impulsivity of participants and their motivation to behave securely, influence decisions made [29], [30].

Table 4: Classification of screenshots according to the used criteria

Alternatives $\mathcal{A}$	**Criteria $\mathcal{G}$**					**Number of Criteria**
	Strength	Padlock	Order	Color	Trust	
OCp	+	+	+	+	+	5
OCn	+	−	+	+	+	4
RCp	+	+	−	+	+	4
OWp	+	+	+	−	−	3
RWp	+	+	−	−	−	2
RWn	+	−	+	−	−	2

5.2. Initial Data for MAVT Model of the Experiment

We now create a MAVT model $M = (\mathcal{A}, \mathcal{G}, \mathcal{K}, V_{\mathcal{G}}, n, W)$, so we specify alternatives $\mathcal{A}$, criteria $\mathcal{G}$, criteria scales $\mathcal{K}$, partial value functions $V_{\mathcal{G}}$ for all criteria, normalization function n, and weights W, followed by the Value Function v.

There are always 6 Wi-Fi's one can select from, so there are six alternatives

$$\mathcal{A} = \{a^1, a^2, a^3, a^4, a^5, a^6\}.$$

We assume five criteria:

$$\mathcal{G} = \{\mathcal{G}^{Strength}, \mathcal{G}^{Trust}, \mathcal{G}^{Padlock}, \mathcal{G}^{Color}, \mathcal{G}^{Order}\}.$$

These correspond to the strength of Wi-Fi signal, the trust provision of the Wi-Fi, the presence of padlock symbol for the network, the color in which the Wi-Fi network name is displayed, and the order position at which network is displayed to the decision maker, respectively.

Each criterion is associated with a scale of decreasing/increasing criteria values: $\mathcal{K} = \{\mathcal{K}^{Strength}, \mathcal{K}^{Trust}, \mathcal{K}^{Padlock}, \mathcal{K}^{Color}, \mathcal{K}^{Order}\}$.

$$\mathcal{K}^{Strength} = \{0, 1, 2, 3, 4\},$$

with strength of signal varying from 0 to 4 bars.

$$\mathcal{K}^{Trust} = \{`Trusted', `Secured', `Open'\},$$

where 'Trusted' indicates that WEP and that the network is in the whitelist of the user; 'Secured' indicates WEP only; and 'Open' indicates neither of the two previous cases.

$$\mathcal{K}^{Padlock} = \{`Present', `Absent'\},$$

indicating whether padlock is used or not.

$$\mathcal{K}^{Color} = \{`Green', `Amber', `White', `Red'\},$$

and it correlates with Trust, as follows: '$Trusted$' is '$Green$'; '$Secured$' is '$Amber$' and '$Open$' is 'Red'. '$White$' refers to a neutral color used for displaying names of Wi-Fi's when no color scheme is applied.

$$\mathcal{K}^{Order} = \{1, 2, 3, 4, 5, 6\}$$

represents the place among the listed networks, with 1 implying place at the top and 6 at the bottom of the list. The normalization function of the form (7) is used here, so all normalized values are between 0 and 1.

When modeling decision making with MAVT, preferences can be provided by (or elicited from) decision makers in a form of criteria weights although this will not be straightforward because of wide range of possible semantics of weights [7]. However, we have data on previous choices of participants available and can use that to estimate weights. We use an ordinal regression approach popular in machine learning literature to extract weights from decisions made by participants of the experiment. Using clustering techniques, we grouped individuals with similar weights in clusters, details will be forthcoming in [31].

Using the experimental data for the OCp screenshot, 7 groups of decision makers were identified, each with different weight vector, see Table 5. For instance, the group $\mathcal{C}_1$ has weights $w_{\mathcal{C}_1} = (0.0089, 0.0793, 0.8064, 0.0463, 0.0590)$. Ignoring the weight values of 1 for the moment, one can see immediately from the high weights in the 'Padlock' criterion column that for many groups the padlock is an important element in their decision making. 'Strength' is also particularly important for some groups (groups $\mathcal{C}_3$ and $\mathcal{C}_4$). The groups $\mathcal{C}_6$ and $\mathcal{C}_7$ represent two extremes where only one criterion is relevant to the decision makers: 'Strength' for $\mathcal{C}_6$ and 'Padlock' for $\mathcal{C}_7$.

Table 5: Groups of decision makers with criteria weights extracted from results for the OCp screenshot.

Clusters $\mathcal{C}$	**Criteria Weights** w				
	Strength	Trust	Padlock	Color	Order
$w_{\mathcal{C}_1}$	0.0089	0.0793	0.8064	0.0463	0.0590
$w_{\mathcal{C}_2}$	0.0973	0.0629	0.7337	0.0433	0.0628
$w_{\mathcal{C}_3}$	0.3199	0	0.6801	0	0
$w_{\mathcal{C}_4}$	0.5566	0.0889	0.1796	0.0618	0.1131
$w_{\mathcal{C}_5}$	0.0858	0	0.9142	0	0
$w_{\mathcal{C}_6}$	1	0	0	0	0
$w_{\mathcal{C}_7}$	0	0	1	0	0

5.3. Analysis of Influence Power per Criterion

In the first part of the analysis we investigate influence power per individual criterion: whether we can change choices of decision makers by modifying one criterion at a time. In the first place, we are interested in studying whether it is possible to influence choices of decision makers by adjusting values on modifiable criteria only: 'Color' $\mathcal{M} = \{'Color'\}$ or 'Order' $\mathcal{M} = \{'Order'\}$. Then we take a step further and check whether other criteria (non-modifiable) would be more influential if we find a way to modify them by considering $\mathcal{M} = \{'Strength'\}$, $\mathcal{M} = \{'Trust'\}$, $\mathcal{M} = \{'Padlock'\}$, $\mathcal{M} = \{'Color'\}$, $\mathcal{M} = \{'Order'\}$. Intuitively, we expect that criteria with highest weights for decision makers are most effective in changing choices.

We now focus on the second group $\mathcal{C}_2$, with criteria weights $w_2 = (0.0973, 0.0629, 0.7337, 0.0433, 0.0628)$, see second row of Table 5. Tables 6–9 display the influence power required to change the selection of the decision maker, providing results for four criteria. The tables should be understood as follows. Each row represents the network for which we want to improve its current rank towards rank of the network in the column: Rank 1 if the network is the user's first choice, rank 2 if it is its second choice, etc. This is known from the experiment data, because users were asked to rank the six networks displayed on a screenshot, i. e. to identify which one they would select first, second, up to sixth. The columns represent the desired improvement of the rank. The values within the space then show the influence power IP, where ∞ means that the influence is not achievable. So, if we consider criterion

Table 6: Cost of influence with 'Strength' criterion for the decision maker from group $\mathcal{C}_2$ with criteria weights $w^{\mathcal{C}_2} = (0.0973, 0.0629, 0.7337, 0.0433, 0.0628)$ for OCp screenshot

Alternatives $\mathcal{A}$	**Alternatives $\mathcal{A}$**					
	'6v2l'	'2xu5'	'71xk'	'dd47'	'7bc2'	'37hh'
'6v2l'	0					
'2xu5'	1	0				
'71xk'	2	1	0			
'dd47'	∞	2	1	0		
'7bc2'	∞	∞	∞	∞	0	
'37hh'	∞	∞	∞	∞	2	0

'Strength' in Table 6, then to move to first position the network that is ranked as second choice ('2xu5'), we need the following influence power $\mathsf{IP} = 1$. To move the worst ranked (6-th) network ('37hh') to first is not possible (∞ for the element in 6-th row and 1-st column in Table 6).

The strong IP for 'Padlock' is obvious from Table 8, where there are no ∞. That is, using only the 'Padlock' criterion, it is possible to make participants of the second group $\mathcal{C}_2$ choose any desired network. For the alternatives ranked as second, third or fourth, $\mathsf{IP} = 1$. This means that simply adding/removing the padlock symbol to one of these alternatives or adding/removing it from the alternative with rank above is enough to improve rank of the alternative. For the alternatives ranked fifth or sixth ranks two changes are needed: adding the padlock to one network and removing the padlock from another. In a similar way analysis for other criteria can be done (see Tables 6, 7, 8 for 'Strength', 'Trust' and 'Order' criteria, respectively). We note that the 'Color' criterion is completely ineffective and we did not present it here. In general, with criteria other than 'Padlock' the influencing possibilities are limited.

5.4. Analysis of Influence Power per Group of Participants and per Screenshot

In the second part of the analysis we want to see how influence power varies over groups of decision makers with different preferences and look at average influence power needed for changing choices of decision makers across all groups for various screenshots. By assuming all criteria as modifiable at the same time $\mathcal{M} = \{'Strength', 'Trust', 'Padlock', 'Color', 'Order'\}$

Table 7: Cost of influence with 'Trust' criterion for the decision maker from group $\mathcal{C}_2$ with criteria weights $w^{\mathcal{C}_2} = (0.0973, 0.0629, 0.7337, 0.0433, 0.0628)$ for OCp screenshot

Alternatives $\mathcal{A}$	**Alternatives $\mathcal{A}$**					
	'6v2l'	'2xu5'	'71xk'	'dd47'	'7bc2'	'37hh'
'6v2l'	0					
'2xu5'	3	0				
'71xk'	∞	1	0			
'dd47'	∞	∞	3	0		
'7bc2'	∞	∞	∞	∞	0	
'37hh'	∞	∞	∞	∞	∞	0

Table 8: Cost of influence with 'Padlock' criterion for the decision maker from group $\mathcal{C}_2$ with criteria weights $w^{\mathcal{C}_2} = (0.0973, 0.0629, 0.7337, 0.0433, 0.0628)$ for OCp screenshot

Alternatives $\mathcal{A}$	**Alternatives $\mathcal{A}$**					
	'6v2l'	'2xu5'	'71xk'	'dd47'	'7bc2'	'37hh'
'6v2l'	0					
'2xu5'	1	0				
'71xk'	1	1	0			
'dd47'	1	1	1	0		
'7bc2'	2	2	2	1	0	
'37hh'	2	2	2	2	1	0

Table 9: Cost of influence with 'Order' criterion for the decision maker from group $\mathcal{C}_2$ with criteria weights $w^{\mathcal{C}_2} = (0.0973, 0.0629, 0.7337, 0.0433, 0.0628)$ for OCp screenshot

Alternatives $\mathcal{A}$	**Alternatives $\mathcal{A}$**					
	'6v2l'	'2xu5'	'71xk'	'dd47'	'7bc2'	'37hh'
'6v2l'	0					
'2xu5'	7	0				
'71xk'	∞	2	0			
'dd47'	∞	∞	7	0		
'7bc2'	∞	∞	∞	∞	0	
'37hh'	∞	∞	∞	∞	∞	0

Table 10: Relative influence power RIP for seven identified groups for 'OCp' screenshot. Lower RIP values suggest less effort needed for successful influencing.

Clusters $\mathcal{C}$	$w_{\mathcal{C}_1}$	$w_{\mathcal{C}_2}$	$w_{\mathcal{C}_3}$	$w_{\mathcal{C}_4}$	$w_{\mathcal{C}_5}$	$w_{\mathcal{C}_6}$	$w_{\mathcal{C}_7}$
RIP **(in %)**	33.33	27.67	26.00	14.00	26.67	24.67	23.33

Table 11: Average relative influence power RIP for all screenshots. Lower RIP values suggest less effort needed for successful influencing. The result indicates that exposing the decision maker to all criteria modifications is not effective (scenarios based on screenshots with less criteria have lower RIP values).

Screenshots $\mathcal{C}$	OCp	OCn	RCp	OWp	RWp	RWn
RIP **(in %)**	23.38	17.33	20.56	17.33	12.33	13.28

we study how big is the fraction of total influence power needed for changing choices of decision makers in groups with different preferences (criteria weights). Even though it is not always realistic for the influencer to have such ability/desire to modify all criteria, here it is used to identify groups of decision makers that are more susceptible to being influenced.

To compute relative influence power RIP, analysis for each screenshot for all groups of participants is performed. We do a separate calculation for each screenshot, since different screenshots pertain to a different set of criteria. Pairs of alternatives at each screenshot are compared and the effort (number of operations to make the decision maker change his/her selection) is computed. For instance, for the OCp screenshot all five criteria are available, see Figure 1 (c). To change the decision from the alternative with all the lowest values on all criteria ('7bc2') to the alternative with all the highest criteria values ('6v2l') the maximal number of criteria changes (altogether 24 operations) is needed. This then is the denominator for RIP. Note also that different criteria have different scales, for which we corrected by assuming that each criterion contributes one-fifth to the denominator in RIP.

The results of analysis of relative influence power are presented in two tables: the results for different clusters of decision makers obtained for 'OCp' screenshot are presented in Table 10, and overall average relative influence power across all screenshots is presented in Table 11. Taking into account

only criteria with non-zero weights the average relative influence power (i. e. the average percentage of changes) needed for all participants and for participants within each group of criteria weights for each screenshot are computed. Analysis of average values in Table 11 confirms our assumption that if there are more criteria involved in alternatives evaluations more effort is required to make a decision maker change his/her selection. In terms of RIP this shows up as follows: the screenshots with lower number of criteria have lower average values for RIP. It seems justified to conclude that if less criteria are available to modify, modification will be more effective, possibly because there is less interference of other artifacts visible to the decision maker.

6. Discussion

Since human minds and abilities are limited, when facing complex choices decision makers tend to simplify problems in order to process and solve them approximately. The first two open issues discussed here are related to such simplification strategies people use to make decisions involving trade-offs. We also demonstrate influence population of decision makers, which may be done in a way similar to influencing individual decision makers.

6.1. Intuitive Strategies for Decision Making Involving Trade-offs

Simplification strategies, also called heuristics, are often applied intuitively. They make us smart and often help us finding fast and frugal solutions as shown with multiple examples in [32]. However, they may also result in unexpected and irrational decisions as discussed in [33].

Applying heuristics is also related to work of our automatic system (which Kahneman calls System 1 [13]) that searches for fast intuitive solutions without using much of cognitive resources, contrary to our controlled system (or System 2) that takes a more deliberate approach but also requires more cognitive resources to make decisions based on some reflexion.

When dealing with alternatives evaluated on several conflicting criteria and representing trade-offs between them, two types of intuitive strategies can be differentiated: compensatory and non-compensatory. In compensatory approaches bad performance of alternatives on some criteria can be compensated by good performance on other ones. In non-compensatory approaches such compensation is not acceptable [34]. In some sense non-compensatory strategies are easier for decision makers, since they have to consider comparison of alternatives on each criterion separately.

Compensatory strategies requires more cognitive effort of decision makers, since they have to compare trade-offs between different criteria in order to make a final choice. Among compensatory approaches, zero-sum heuristic or compensatory inference, derives from the game theory idea of balancing gains and losses of all players. This concept is extended to economics with efficient markets [11], where it is assumed that products with high prices have high values. Based on this approach, in situation of uncertainty and absence of information on some of criteria, people derive missing criteria values from the values of observable criteria assuming correlation between them. For instance, quality of a product is deduced from its price, or quality from a brand name.

It was shown in [35] that similarly users derive unobservable security quality of Wi-Fi from its observable convenience quality. This research is particularly interesting in the context of the present work, since it suggests that a majority of people perceive security of Wi-Fi's wrongly: Using compensatory inference people underestimate security quality of a Wi-Fi with both high security and high convenience, and overestimate its convenience, when comparing it to a Wi-Fi with both lower security and lower convenience. It was demonstrated experimentally that such wrongly perceived qualities of both security and convenience affect choices of Wi-Fi's by decision makers, and that the level of users' technology knowledge had no significant impact on their choice.

6.2. Strategies to Influence Decision Making Involving Trade-offs

In the context of the present work, particularly interesting are results on influencing decision making involving multiple criteria presented by Sedikides et al. in [36], where contextual and procedural determinants were studied (for the case study of a partner selection). As a contextual determinant of choice, the asymmetric dominance effect [36] (also called 'decoy' [33]) was investigated, which refers to the phenomenon of changing preference from one alternative to another after introduction of a third alternative, which is very similar to (although dominated by) the alternative towards which the decision maker is nudged, but non-dominated with respect to the other alternative. Although the third alternative is not likely to be selected as it is inferior to one of the alternatives, it has a role of highlighter of the alternative which dominates it. Ease of comparison of this newly introduced alternative to one of the alternatives makes the last one more salient and increases its chances to be selected when compared to another non-dominated alternative.

For the Wi-Fi choice example introduced in Section 4.1 and presented in Table 1 in order to nudge decision maker towards selecting a secure alternative with weak signal s, a *decoy* alternative d may be introduced such that $v = [s \mapsto (v_1, v_2, v_3), d \mapsto (v_1, v_7, v_3), f \mapsto (v_4, v_5, v_6)]$, where $v_7 < v_2$. For instance, $v = [s \mapsto (1, 1, N), d \mapsto (1, 0, N), f \mapsto (0, 2, N)]$, meaning that the decoy alternative has the same values on the trust and color criteria, but is worse on the strength of signal criterion when compared to (and consequently is dominated by) the alternative s. For instance, in the Wi-Fi example the decoy alternative d has worse value of trust when compared to the alternative s. This difference indicates to the decision makers how much they are losing by not selecting the alternative s. On the other hand, there is no such highlighter for the alternative f, which is very different from both s and d even though non-dominated with respect to both.

The explanation of why decoy works can be found in another research on prospect theory by Kahneman and Tversky [12], which shows that regret from losing is much stronger then the joy of winning, and people would rather not lose than win. The decoy alternative shows to the decision makers exactly how much they are loosing by choosing an alternative other than the alternative similar to the decoy alternative. Obviously, the effect can be neutralized by introducing another decoy alternative, which is similar but is a bit worse than the second alternative, the alternative f with strongest signal in the Wi-Fi example.

6.3. Influencing Population

In all previous sections, we have considered a deterministic decision maker by default. To allow modeling groups of decision makers rather than single users, we may consider a *probabilistic* decision maker. We model this aspect by considering a probability distribution over weights, such that given a weight function w, $\psi(w)$ represents the probability of w. From a statistical point of view, $\psi(w)$ represents the percentage of the population with the weight distribution w.

The policy of the entire population can therefore be defined as given in MAVT model $M = (\mathcal{A}, \mathcal{G}, \mathcal{K}, V_{\mathcal{G}}, n, \psi(w))$:

$$\pi(\psi, v, a) = \sum_{w \in W} \{\psi(w) \mid \pi(w, v,) = a\}. \tag{8}$$

For influencing a population of decision makers, an influencer needs to look for an alternative (or subset of alternatives) with highest impact and

a subset of modifiable criteria that makes this alternative (preferred by the influencer) to be selected by the majority of population.

$$\mathsf{opt}(w_c, v_c, \psi_u, v_u) = \arg \max_{v' \in P_{\mathcal{M}}(v_u)} \sum_{a \in \mathcal{A}} \rho(a, w_c, v_c)\pi(\psi_u, v'_u, a).$$

As an example of population modeling, we can consider examples of three types of decision makers with the same criteria evaluation functions $v_u = [s \mapsto (1, 1, N), f \mapsto (1, 2, N)]$, but different criteria weights $w_1 = (0.3; 0.5; 0.2)$, $w_2 = (0; 0.8; 0.2)$, and $w_3 = (0.8; 0; 0.2)$. Let us also consider a probability distribution ψ such that $\psi(w_1) = \psi(w_2) = \psi(w_3) = 1/3$. We can calculate that $\pi(w_1, v_u) = f$, $\pi(w_2, v_u) = f$, $\pi(w_3, v_u) = s$, and therefore, following Equation (8), we have $\pi(f, v, \psi) = 2/3$ and $\pi(s, v, \psi) = 1/3$.

If the influencer wants to shift choices of a population of users, he/she may consider similar strategy as one proposed for influencing choice of individual decision makers, but taking into account weights of different groups of users.

7. Conclusions and Outlook

In this paper we proposed an operational model for influencing human decision making in security context. In particular, we introduced influence within multi-attribute value theory. The set-up for modeling influencing requires two parties, a decision maker and an influencer, each of them with their own MAVT optimization model. The influencer's attempts to influence a decision are modeled through changes in the MAVT model of the decision maker and to express such changes we introduced the concept of modifiable criteria. We also introduced the concept of influence power, which quantifies the effort to successfully influence the decision maker.

The application of the model using data collected in a study of public Wi-Fi network selection resulted in a number of interesting insights. First, we showed it is possible using data from actual experiments to express influencing in a MAVT model with modifiable criteria. In addition, using influence power, we were able to identify which optimization criterion is most amenable to influencing, by calculating for which criteria successful influencing requires the least effort. For instance, in the Wi-Fi example, we showed that adding the padlock symbol to a displayed network name was by far the most powerful influencer for the largest group of participants. An additional important aspect when influencing is the fact that (groups of) people are ideally targeted in bespoke manner. In the Wi-Fi study, several 'clusters' of decision

makers were identified with often very different criteria determining their decision. Finally, the analysis also indicated that influence power decreases when increasingly many options are available to the influencer, possibly because the affect of a single change in the display is less pronounced when other display features compete for attention from the decision maker. The latter two items would need to be researched in more detail.

In general, formal modeling and optimization models such as these in this paper should ideally be integrated more intimately with the design of influencing attempts. Methods for designing choice architectures (i.e., the way choices are presented to the decision maker) can use the models to identify which elements in the choice architecture can be expected to be most powerful in influencing decisions. In addition, the work presented in this paper provides an approach to personalizing influencing attempts by calculating influence power depending on the characteristics of different groups of decision makers. From our study it became clear that personalization, if practicable, potentially greatly enhances the likelihood of success of influencing.

Acknowledgements

The work presented in this paper was funded in part through Choice Architecture for Information Security (EP/K006568/1) from Engineering and Physical Sciences Research Council (EPSRC), UK, and Government Communications Headquarters (GCHQ), UK, as a part of the Research Institute in Science of Security. We gratefully acknowledge the support and contribution of our colleagues in the project, from Newcastle University: James Turland and Thomas Groß, and from Northumbria University: Debora Jeske, Lynne Coventry, Pam Briggs and Christopher Laing. Additional support was provided through University of Illinois at Urbana-Champaign's NSA Lablet in the Science of Security. We are also thankful to unknown reviewers that helped to improve this work.

References

[1] J. Clarke, M. G. Hidalgo, A. Lioy, M. Petkovic, C. Vishik, J. Ward, Consumerization of IT: Top risks and opportunities, Tech. rep., European Network and Information Security Agency (ENISA) (2012).

[2] R. H. Thaler, C. R. Sunstein, Nudge: Improving decisions about health, wealth, and happiness, Yale University Press, New Haven, CT, USA, 2008.

[3] Applying behavioural insights to reduce fraud, error and debt, Policy paper: Transforming government services to make them more efficient and effective for users, Cabinet Office, Behavioural Insights Team, UK (February 2012).

[4] A. S. Hanks, D. R. Just, B. Wansink, Trigger foods: The influence of irrelevant alternatives in school lunchrooms, Agricultural and Resource Economics Review 41 (1) (2012) 114–123.

[5] C. Holley, Helpdesk report: Use of behavioural economics in development interventions, Helpdesk reports by the British Governments Department for International Development, human development group, Human Development Resource Center, UK (2012).

[6] C. Morisset, I. Yevseyeva, T. Groß, A. van Moorsel, A formal model for soft enforcement: Influencing the decision-maker, in: S. Mauw, C. Jensen (Eds.), Security and Trust Management, Vol. 8743 of Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 113–128.

[7] V. Belton, T. Stewart, Multiple criteria decision analysis: An integrated approach, Kluwer Academic Publishers, Dordrecht, 2002.

[8] R. Keeney, H. Raiffa, Decisions with multiple objectives: Preferences and value tradeoffs, J. Wiley, New York, 1976.

[9] J. Turland, L. Coventry, D. Jeske, P. Briggs, A. van Moorsel, Nudging towards security: Developing an application for wireless network selection for Android phones, in: Proceedings of the 2015 British HCI Conference, British HCI '15, ACM, 2015, pp. 193–201.

[10] I. Yevseyeva, C. Morisset, T. Groß, A. van Moorsel, A decision making model of influencing behavior in information security, in: A. Horvth, K. Wolter (Eds.), Computer Performance Engineering, Vol. 8721 of Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 194–208.

[11] A. Chernev, Jack of all trades or master of one? Product differentiation and compensatory reasoning in consumer choice, Journal of Consumer Research 33 (4) (2007) 430–444.

[12] D. Kahneman, A. Tversky, Prospect theory: An analysis of decision under risk, Econometrica 47 (2) (1979) 263–291.

[13] D. Kahneman, Thinking, fast and slow, Farrar, Straus & Giroux, New York, 2011.

[14] E. K. Choe, J. Jung, B. Lee, K. Fisher, Nudging people away from privacy-invasive mobile apps through visual framing, in: INTERACT (3), Vol. 8119 of LNCS, Springer, 2013, pp. 74–91.

[15] L. Coventry, P. Briggs, D. Jeske, A. van Moorsel, SCENE: A structured means for creating and evaluating behavioral nudges in a cyber security environment, in: A. Marcus (Ed.), Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience, Vol. 8517 of Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 229–239.

[16] R. Balebako, P. G. Leon, H. Almuhimedi, P. G. Kelley, J. Mugan, A. Acquisti, L. F. Cranor, N. Sadeh, Nudging users towards privacy on mobile devices, in: Workshop on Persuasion, Influence, Nudge and Coercion Through Mobile Devices (PINC at CHI-11), 2011.

[17] P. Dolan, M. Hallsworth, D. Halpern, D. King, I. V. R. Metcalfe, Influencing behaviour: The MINDSPACE way, Journal of Economic Psychology 33 (2) (2012) 264–277.

[18] A. Beautement, R. Coles, J. Griffin, C. Ioannidis, B. Monahan, D. Pym, A. Sasse, M. Wonham, Modelling the human and technological costs and benefits of usb memory stick security, in: M. Johnson (Ed.), Managing Information Risk and the Economics of Security, Springer US, 2009, pp. 141–163.

[19] D. Eskins, W. H. Sanders, The multiple-asymmetric-utility system model: A framework for modeling cyber-human systems, in: Proceedings of the 2011 Eighth International Conference on Quantitative Evaluation of SysTems, QEST '11, IEEE Computer Society, Washington, DC, USA, 2011, pp. 233–242.

[20] N. Thomas, M. Harrison, Y. Zhao, X. Chen, Formal performance modelling: From protocols to people, in: M. Tribastone, S. Gilmore (Eds.), Computer Performance Engineering, Vol. 7587 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2013, pp. 127–139.

[21] R. Bellman, A Markovian decision process, Journal of Mathematics and Mechanics 6 (4) (1957) 679–684.

[22] A. Y. Ng, S. J. Russell, Algorithms for inverse reinforcement learning, in: Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000, pp. 663–670.

[23] K. Miettinen, Nonlinear Multiobjective Optimization, Kluwer Academic Publishers, Dordrecht, 1999.

[24] R. Hoes, T. Basten, C.-K. Tham, M. Geilen, H. Corporaal, Quality-of-service trade-off analysis for wireless sensor networks, Performance Evaluation 66 (35) (2009) 191 – 208.

[25] K. Avrachenkov, U. Ayesta, A. Piunovskiy, Convergence of trajectories and optimal buffer sizing for AIMD congestion control, Performance Evaluation 67 (7) (2010) 501 – 527.

[26] C. Morisset, T. Groß, A. van Moorsel, I. Yevseyeva, Nudging for quantitative access control systems, in: T. Tryfonas, I. Askoxylakis (Eds.), Human Aspects of Information Security, Privacy, and Trust, Vol. 8533 of Lecture Notes in Computer Science, Springer International Publishing, 2014, pp. 340–351.

[27] A. Ferreira, J.-L. Huynen, V. Koenig, G. Lenzini, S. Rivas, Socio-technical study on the effect of trust and context when choosing WiFi names, in: STM, Vol. 8203 of LNCS, Springer Berlin Heidelberg, 2013, pp. 131–143.

[28] G. Farnham, K. Leune, Tools and standards for cyber threat intelligence projects, Technical report, SANS Institute (2013).

[29] D. Jeske, L. Coventry, P. Briggs, Decision justifications for wireless network selection, in: 2014 Workshop on Socio-Technical Aspects in Security and Trust (STAST), 2014, pp. 1–7.

[30] D. Jeske, L. Coventry, P. Briggs, A. van Moorsel, Nudging whom how: IT proficiency, impulse control and secure behavior, in: Personalizing Behavior Change Technologies CHI Workshop, Toronto, Canada, 2014.

[31] I. Yevseyeva, C. Morisset, A. van Moorsel, Predicting security choices with a multi-criteria model extracted by ordinal regression: A case study for connecting to a public Wi-Fi, 2015, (in preparation).

[32] G. Gigerenzer, P. M. Todd, the ABC Research Group, Simple heuristics that make us smart, 1st Edition, Oxford University Press, Oxford, UK, 1999.

[33] D. Ariely, Predictably irrational: The hidden forces that shape our decisions, 2nd Edition, HarperCollins Publishers, Hammersmith, London, UK, 2009.

[34] P. Goodwin, G. Wright, Decision analysis for management judgment, 4th Edition, J. Wiley, 2009.

[35] B. C. Kim, Y. W. Park, Security versus convenience? An experimental study of user misperceptions of wireless Internet service quality, Decision Support Systems 53 (1) (2012) 1–11.

[36] C. Sedikides, D. Ariely, N. Olsen, Contextual and procedural determinants of partner selection: Of asymmetric dominance and prominence, Social Cognition. Special Issue: Social Cognition and Relationships 17 (1999) 118–139.