D’utilisable à véritablement engageant : Involved, un système de gestion énergétique persuasif pour les bâtiments

From Usable to Incentive-Building Energy Management Systems

Amr Alzhouri Alyafi[1,3], Van Bao Nguyen[2], Yann Laurillau[2], Patrick Reignier[1], Stéphane Ploix[3], Gaëlle Calvary[2], Joëlle Coutaz[2], Monalisa Pal[4], Jean Philippe Guilbaud[2]

[1] Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP*, LIG, 38000 Grenoble, France
[2] Univ. Grenoble Alpes, CNRS, Grenoble INP*, LIG, 38000 Grenoble, France
[3] Univ. Grenoble Alpes, CNRS, Grenoble INP*, G-SCOP, 38000 Grenoble, France
[4] Machine Intelligence Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Kolkata - 700108, India

* Institute of Engineering Univ. Grenoble Alpes

ABSTRACT. Reducing energy consumption is an individual and collective challenge that requires people to be proactive and fully involved. However, most approaches and solutions to this problem promote automated and autonomous systems that take full control of the decisions. Although these systems relieve users from setting the temperature in offices and homes, the level of thermal comfort chosen by the technology may not meet the individuals' requirements. Automating control is an obvious solution but it keeps the human out of the loop. In this paper, an alternative approach is promoted based on system–human cooperation that allows people to make the final decision. However, to make sound decisions, users need to understand the system functioning and rationale and they need to be convinced and motivated to change their habits with regard to energy consumption. Our approach is based on an e-coach system, called Involved, which provides users with contextual explanations along with a user interface designed to persuade the user to stay involved. The results of this research are the development of an early prototype that provides end-users with a 24-hour plan of recommended actions along with contextual explanations that justify each action. This plan satisfies the inhabitant’s preferred compromise between the thermal comfort, air quality and financial cost specified by the inhabitant (i.e. the user) using a novel interaction technique, the Trade-Off-Pareto sliders. Users can edit the plan (e.g. suppress an action), skip some actions, perform additional actions, or even change their preferred compromise, and be informed in real time of the consequences on energy consumption and comfort.

RÉSUMÉ. Réduire la consommation énergétique est un défi individuel et collectif qui exige que les personnes soient proactives et pleinement impliquées. La plupart des approches et des solutions actuelles reposent sur des systèmes automatisés et autonomes prenant le contrôle total des décisions. Bien que ces systèmes libèrent les occupants du réglage de la température dans leur bureau ou leur maison, le niveau de confort thermique choisi par la technologie peut ne pas satisfaire leurs exigences. L'automatisation du contrôle est une solution évidente, mais elle écarte l'humain du processus. Nous proposons une approche alternative reposant sur la coopération entre un gestionnaire énergétique et les occupants leur permettant de prendre la décision finale. Cependant, pour prendre de bonnes décisions, les occupants doivent comprendre le fonctionnement et la logique du système et être convaincus et motivés pour changer leurs habitudes en matière de consommation énergétique. Notre approche propose un assistant virtuel qui fournit aux occupants des explications contextuelles ainsi qu'une interaction homme-machine conçue pour persuader et maintenir l'utilisateur impliqué.

KEYWORDS. Contextual explanations, incentive system, persuasive technology, system intelligibility, human–computer interaction, smart building, energy management.

MOTS-CLÉS. Explications contextuelles, motivation, technologie persuasive, intelligibilité, interaction homme-machine, gestion énergétique.

1. Introduction

According to statistics from the French Ministry of Environment, Energy and Sea published in 2017 [1], 45% of the total energy consumption in France can be attributed to the residential and tertiary sectors. In the global context of control and reduction of energy consumption, the residential sector is thus an important domain to consider. With poorly insulated houses, regardless of what the inhabitant does, energy is wasted. On the other hand, modern homes are increasingly efficient. They are equipped with complex heating and cooling devices and they are able to produce their own energy. As components of smart grids, modern homes will soon have to provide energy flexibilities to even more complex systems. As this occurs, the main source of energy waste and flexibility will not be the home *per se*, but inhabitants' inappropriate behaviors.

Traditional approaches to complex building management promote automation as a *doing instead of* paradigm (i.e. the system has full control). Using a mathematical model of the thermal characteristics of the house and its devices, optimization approaches based on objective functions typically generate a daily "Pareto optimal plan of action" used as input by an energy manager to control the home. The plan is both "Pareto optimal" and "non-optimal" as it is a compromise between some level of comfort and the amount of energy necessary to reach this level of comfort. From the end-user's point of view, the house acts autonomously. In addition, end-users might not understand the system behavior and may even feel that they have to fight against their home in order to reach their objective. A typical example of this situation is a user who closes the blinds to adapt light conditions for watching TV. As a result, the system automatically switches the lights on, which is inappropriate in the current situation. Moreover, a couple of minutes later, the system may re-open the blinds to heat the room on a sunny winter day! Although the equations that model the home and the objective functions appear to be complete from the point of view of physics, they do not take into account human factors such as intention. As a result, the system has nice properties from the theoretical perspective of energy consumption but may be inappropriate for the inhabitant's daily living.

Rather than *doing instead of*, we promote a *doing together* approach where the inhabitant (i.e. the user) is involved in the final decision and actions. However, sound decisions and actions require users to understand the system functioning and rationale. Users may not be able to answer questions such as "for better thermal comfort and heating performance, shall I shut the kitchen window now? When leaving for work in the morning, shall I switch off the heater and switch it on when I return, or shall I only decrease the temperature set point?"

In this article, we describe Involved, an e-coach that supports end-users in finding optimal compromises between tightly coupled, and possibly conflicting, criteria such as thermal comfort, air quality and financial cost. Based on the specification of user requirements, Involved is able to produce contextual recommendations for actions along with explanations presented in a suitable manner. Using the Involved "what-if" feature, the user can edit the daily plan of recommended actions (such as suppressing the "Close the kitchen window by 10:00 am" recommendation) while being informed of the consequences on energy consumption and comfort. As demonstrated by Karsenty and Brézillon [2], it is reasonable to hypothesize that contextual explanations coupled with the capability to explore the possible configurations in a "what if" manner will allow users to better understand the functioning and rationale of the energy management system, and will be *prompted* to act in a more virtuous manner while remaining *motivated* in the long run. Actually, Involved, as a proof of concept of a persuasive e-coach, demonstrates a new interactor that is dedicated to navigating among conflicting criteria, as well as an explanation engine. The prototype is currently being tested with users.

This article is structured as follows. An overview of the background knowledge related to persuasive interactive systems is presented in the following section. Based on the models and design principles developed for supporting sustainable behavior change, the concept of explanation is presented in detail in section 3. Involved is then described according to two complementary perspectives: the functional perspective with a focus on the generation of contextual explanations in section 4, followed by the interactional perspective in section 5, with a focus on the user interface.

2. Persuasive interactive systems

Persuasive technology refers to “interactive technology that changes a person’s attitudes or behaviors”[3]. According to Ajzen [4], attitude is one determinant for behavior whereas for Festinger [5], behavior and attitude may both be changed for the sake of consistency. Fogg introduced the concept of captology to denote persuasive technology as a research area. Captology does not include computer-mediated persuasion nor non-intentional persuasion (Fogg, 2003). Oinas-Kukkonen and Harjumaa [6] prefer the term Behavior Change Support Systems (BCSS). Contrary to captology, BCSS can play the role of a mediator between two persons. In both cases, coercion and betrayal are excluded from the field for ethical reasons.

Persuasive interactive systems result from applying persuasive technology to the engineering of interactive systems. These systems are expected to sense inappropriate or undesirable behavior, such as smoking, and then enact functions, such as recommendations, that are suitable to support users’ changes. The field draws upon a large body of models and theories such as the technology acceptance model [7] and the Motivation–Opportunity–Ability model [8]. In the following, we will limit the presentation to two seminal models that currently serve as references for the development of interactive persuasive systems: Fogg’s behavior model (FBM) and the Transtheoretical model of change (TTM). We will then summarize the design principles drawn from these models and illustrate the discussion with representative examples of persuasive interactive systems related to energy management and water usage.

2.1. *FBM and TTM: two reference models for persuasive interactive systems*

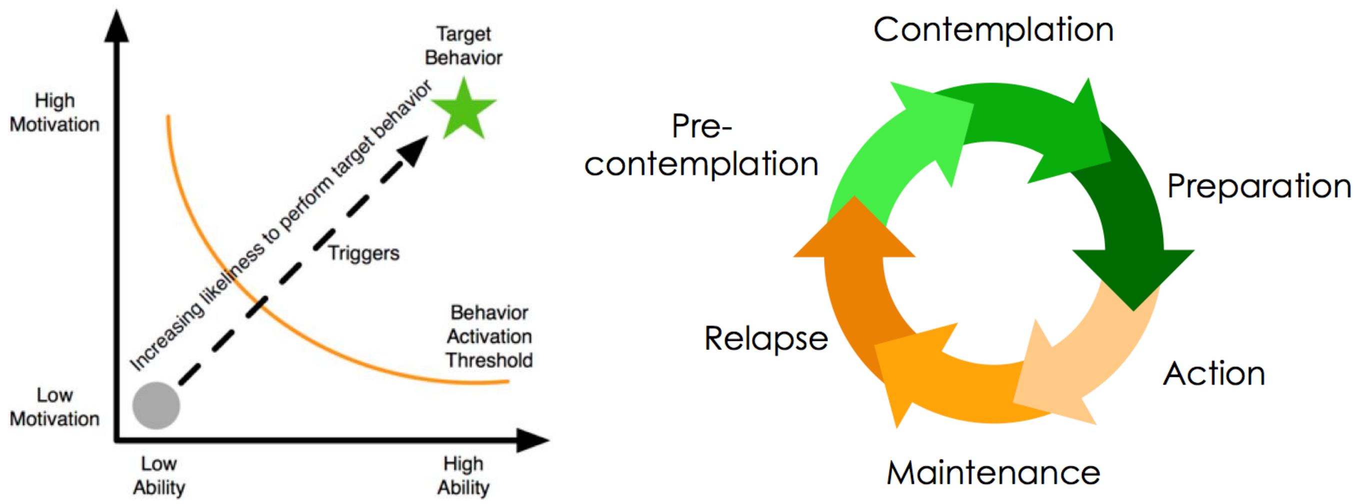


Figure 1. *Fogg's behavior model (left) and the transtheoretical model of change (right)*

Fogg’s behavior model characterizes human behavior according to three dimensions [9]: *motivation* to adopt a particular target behavior, *ability* to perform the target behavior, and *triggers*. As shown in Figure 1 (on the left), motivation and ability define a 2D-space where both human behavior and resistance to change can be characterized. Any trigger received by an individual while they are above the activation threshold in the 2D-space should make this person adopt the behavior. In turn, motivation can evolve along three dimensions: pleasure/pain, hope/fear, and social acceptance/rejection. As learning new skills is not frequently accepted by people, ability can simply rely on the simplification of human behavior.

The transtheoretical model of change (TTM) [10], [11] decomposes behavior change into a six-stage cycle (Figure 1, on the right):

- Pre-contemplation: subjects are not considering the idea of change, perhaps because they are unaware or uninformed or possibly because they are frustrated by a previous failed change attempt. They do not intend to take action in the next six months.

- Contemplation: subjects are aware that they should change behavior, and they consider attempting the change in the next six months. In this stage, they try to get informed about the problem, but they are not ready to take concrete action toward change.

- Preparation: subjects are ready to make a change in the near future (usually measured as the next month), and they are trying to develop a plan to take their first concrete action in the direction of change.

- Action: subjects have taken action and modified their behavior within the past six months.

- Maintenance: subjects try to keep the behavior change, and struggle to prevent relapsing. In the case of failure at this stage, a relapse will occur and regress them to an earlier stage. They will have to restart the progress from one of the first stages.

- Relapse: subjects fail at the target behavioral change.

2.2. *Persuasive design principles*

According to Fogg, technology can play three roles [12]: as a tool, as media, and as a social actor.

As a tool, technology can make activities easier or more efficient to perform. The corresponding design principles are thus “reduction”, “tunneling”, “tailoring”, “suggestion”, “monitoring” and “self-monitoring”, and “conditioning”. In particular:

- Suggestion: persuasion power can be increased by offering suggestions about behavior change. As described in section 5, one of the key features of Involved is its capacity to provide users with recommended action plans along with the appropriate contextual explanations;

- Monitoring and self-monitoring: technology “eliminates the tedium of tracking performance”. As such, it serves as the basis for revealing behavior or for monitoring progress. Mirroring users’ behavior and their impact on comfort level and energy savings is part of the Involved solution;

- Conditioning: positive reinforcement can be used to “transform existing behaviors into habits”.

As media, technology can shape attitudes and behaviors by providing compelling simulated experiences. The corresponding design principles are “cause and effect”, “virtual rehearsal”, “virtual rewards” and “simulations in real-world contexts”. For instance, Fogg [12] defines the “cause and effect” principle as a means to persuade people to make a change as simulation can make observable “the link between cause and effect”. In section 4, we will see how this “cause and effect” principle is applied to the generation of contextual explanations as well as the “what if” feature that allows users to simulate the effects of alternative behavior on energy consumption.

As social actor, technology persuades by giving a variety of social cues that elicit social responses from their human users. The corresponding design principles are “attractiveness”, “similarity”, “praise”, “reciprocity” and “authority”. Among these principles, focusing on mobility, “social comparison” is another related principle [12]: comparison with the performance of others can increase motivation. In its current version, the social dimension has not been investigated for Involved. On the other hand, special attention has been paid to “attractiveness”.

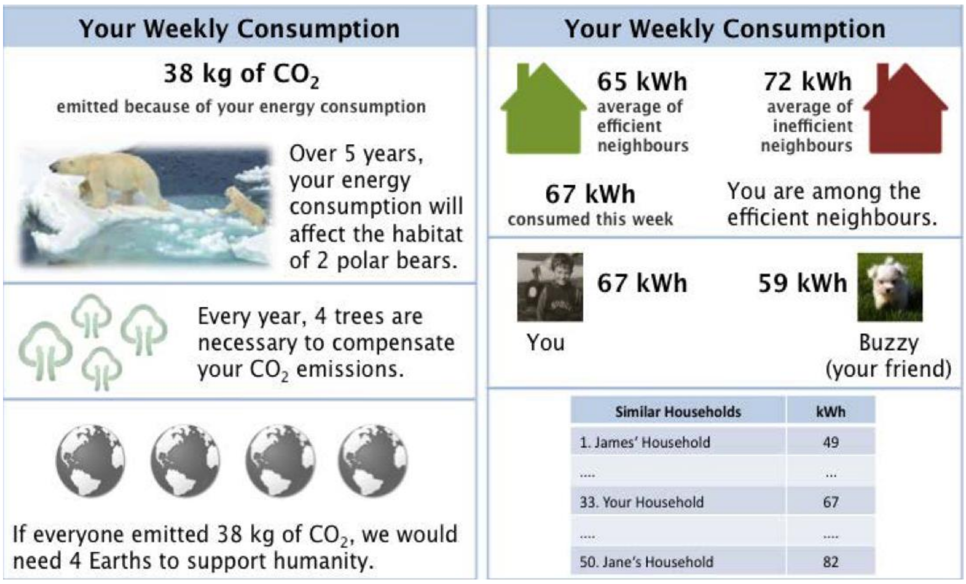
Applied to a case study about water consumption, [13] identifies seven design principles: "value-added design", "automation", "just-in-time prompts", "positive reinforcement", "negative reinforcement", "adaptive interfaces" and "social validation".

In summary, the domain of persuasive technology is an active area of research with the development of a variety of theories, models and design principles. As discussed in sections 4 and 5, a number of these models and principles have been used for the development of Involved. The following section briefly describes a subset of representative persuasive interactive systems for energy and water management.

2.3. *Examples of persuasive interactive systems for energy and water management*

Figure 2. *Power-aware cord (left) and water bot (right)*

The power-aware cord [14] (Figure 2, on the left) uses a electrical cord with lights that make the electricity flow tangible. The water bot [13] (Figure 2, on the right) is designed to encourage users to reduce water consumption by providing immediate feedback about water usage. A device attached to the faucet keeps track of water usage and reminds users to save water by changing the color of the water flow according to water usage.



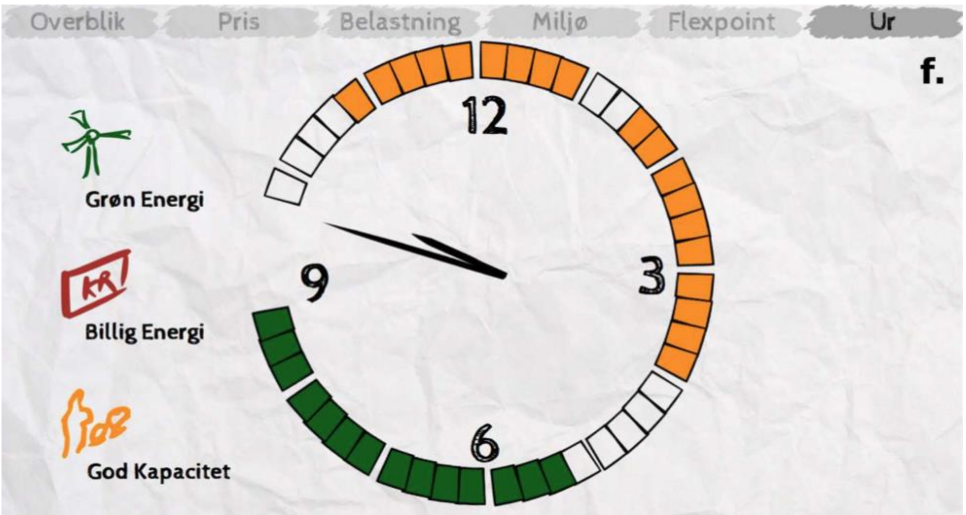


Figure 3. *Possible savings (top) and a clock-view representation of electricity cost (Grøn Energi = green energy; Billig Energi = cheap energy; God Kapacitet = good capacity) (bottom)*

Petkov [15] (Figure 3, at the top) proposes an approach to reduce energy consumption based on possible savings. It presents the financial gain that users can obtain by using less energy. It also highlights the impact of energy consumption on health, resources, and environment. Using a clock-view representation, eForecast [16] indicates the peak hours for a domestic household as well as the green zone when electricity is cheap (Figure 3b).

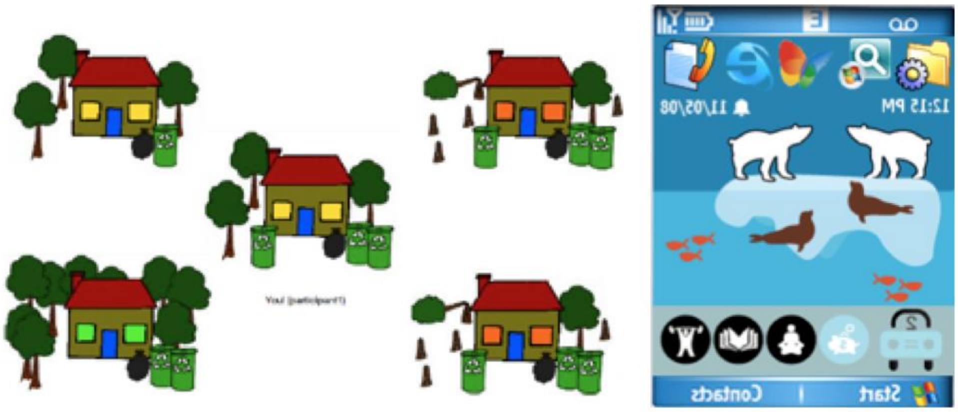


Figure 4. *Elsmore et al.'s user interface (left) and Ubi-green (right)*

Elsmore *et al*. [17] explore human behavior by combining two domains: energy and sobriety. As shown in Figure 4 (on the left), the user interface shows five households where the overall result is conveyed by the color of the windows, while the number of trees and trashcans reflect electricity consumption and waste respectively. Ubi-green [18] uses the metaphor of a polar bear standing on an iceberg to illustrate household consumption (Figure 4, on the right). The objective is to keep the polar bear alive and to improve the ecosystem. Informative art is also considered for integrating "information visualization in the everyday human environment" [19] (e.g. Mondrian's style for a weather display). In the context of physical activities, it has been found [20] that it is motivating and engaging to glance at a display of aesthetic informative art. Also, in the field of persuasive technologies, paintings are investigated to convey emotions about behavior change, such as the Mona Lisa bookshelf [21]. According to Fogg's behavior model (FBM) [9], the informative art approach could increase motivation as it is related to several motivational factors: pleasure factor increased by the aesthetic aspects; hope and fear factors impacted by the utility services such as current time (e.g. hope of being on time, fear of being late) or current weather (e.g. hope of having a sunny day). Furthermore, in terms of operant conditioning, a changing aesthetic could be considered as a reward for our past actions.

According to the TTM [10], [11], it would support the transition between the pre-contemplation stage and the contemplation stage as well as the maintenance stage (i.e. maintaining interest over time).

All of these systems have limitations. First, behavior change is a long-term process rather than a single event [10], [11]. The power-aware-cord and the water bot can only be effective at the TTM "pre-contemplation" stage, as they seek to plant a seed in the user's mind by visualizing energy consumption. Such systems are unlikely to be effective in motivating users to commit to long-lasting virtuous behavior.

The second problem relates to motivational incentive in behavior change. Petkov uses potential savings as a motivational incentive. However, financial incentive does not lead to sustainable behavior change [15]. Moreover, when the cost of energy is low compared to the individual's income, the motivation is unlikely to be effective. As a consequence, e-Forecast is unable to provide a sustainable solution.

Third, the use of social influence and value-centered design [22] in persuasive energy systems is mentioned but still remains scarcely exploited. Elsmore *et al*. explore how community affects individual changes, while Ubi-green considers the intrinsic values [23] of humans as the key factor to raise awareness about energy consumption. However, like the Elsmore's system, Ubi-green mainly focuses on comparison between households. Some participants will consider the system as a competition resulting in a variety of objectives: some will try to win while others may do whatever it takes to not finish last. In Ubi-green, the possibility of raising altruistic values is presented. The fact that some participants may treat it as a game will negatively impact the original aims of the application.

Last but not least, user interaction is limited in all of these systems. They make energy visible, but do not support "what if" interactions. FigureEnergy (Figure 5) [24] moves from user interfaces to user interaction, making it possible for the end-user to play with the system.

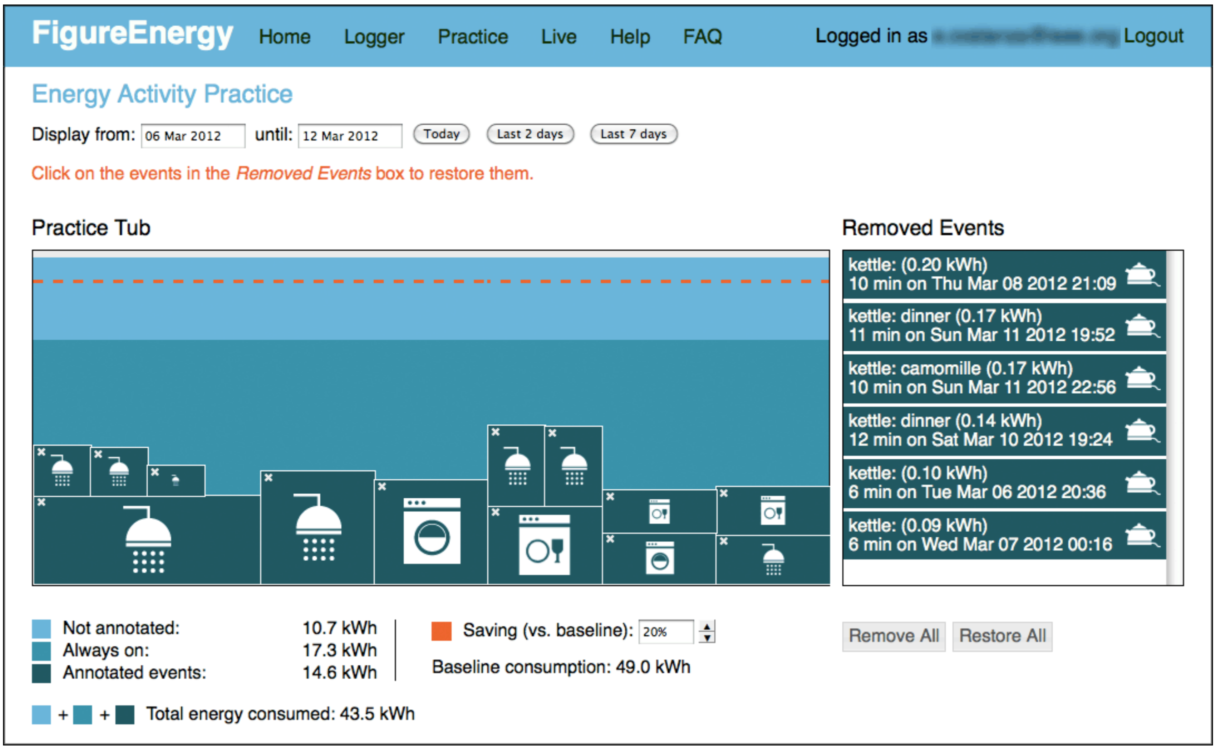


Figure 5. *From user interface to user interaction in Costanza et al. [24]*

3. Explanations and causality

Explanations play different roles and take a variety of forms. Already addressed in the previous section about persuasion, causality is also an important concept involved in the production of explanations. These aspects are discussed successively in this section with a particular focus on the problem of elaborating a formal model of causality from sensory data. This section closes with the qualities that explanations should satisfy.

3.1. *Explanations in knowledge-based systems*

The 80s saw the rise of expert systems (and more generally knowledge-based systems, or KBS) and their adoptions in corporations as part of the information system. Expert systems are usually based on an explicit high-level symbolic knowledge database and an inference engine to exploit them. They are able to answer questions and make decisions in their domains of expertise. Beyond the decision-making aspect, the question of the generation of corresponding explanations was quickly raised and became an important field of study (see [25]–[27] for instance).

Gregor *et al.* [25] stated that explanations in intelligent systems are important for several reasons. An expert user trusting their KBS will need explanations from the system mainly when he or she detects anomalies (when he or she disagrees with its propositions). On the contrary, a novice user will use the explanation capacity of the system more but with a short-term or long-term learning objective. By providing explanations, a system becomes more transparent. It also appears more competent, and trust in an automated systems is related to the user's perception of its competence (see [28] for instance).

Explanations are also required by the users when they will lack some knowledge needed to contribute properly in a problem-solving process. Neither the KBS nor the user can solve the problem alone; they have to co-operate. Users have their own expertise that may differ from the expertise of the KBS, and they know the context in which the problem occurs. "Computational technology should be used, not to make or recommend solutions, but to help users in the process of reaching their decision" [2]. This is what Woods *et al.* called a "Joint Cognitive System" [29]. The system and the user are engaged in a cooperation process that relies on explanations.

Karsenty and Brézillon have studied cooperative dialogs in natural working conditions [2]. Two different types of dialogs were considered: validation dialogs and design dialogs. Validation dialogs are between an expert and a user. The expert proposes a solution that is validated or modified by user. The expert and the user have complementary knowledge: the expert has the technical knowledge and the user has the domain expertise. In design dialogs, a first solution is not proposed by one of the experts but the preliminary solution is proposed from the problem definition by a close cooperation between both, each one bringing their technical expertise.

Context plays a central role in the human–system cooperation process when solving a problem. It is an important element to produce more pertinent explanations and thereby enhance cooperation. Defining context has been a challenging problem for the last 20 years (see [30] for an example of an early survey). Its definition may even depend on the context [31]. In the ubiquitous computing domain, Dey has proposed a definition focusing on human–machine interaction: "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves" [32]. Context can be considered from both a representational or interactional point of view [31], [33]. Context as representation is information. It can be separated from activity. Contextual elements can be selected and described before the interaction occurs and are not modified by the interaction: activity happens within the context [33]. Context as interaction is a relation between activities or objects. It only exists during the interaction (task or problem-solving) and evolves dynamically.

As we mentioned in the introduction, building a fully automated energy manager (*doing instead of*) may be inappropriate for users. We are in the same situation as described by Woods *et al*. or Karsenty *et al*.: the system has some knowledge and competence to solve the problem, with the notable difference that this knowledge is numerical (physical knowledge, optimization capabilities, etc.) and not symbolic, and it is therefore more intelligible than in the studies cited above. The user has their own complementary knowledge (like their own perception of comfort, their intentions, etc). Neither the system nor the user can individually solve the whole problem but they have to cooperate (joint cognitive system). This cooperation corresponds to what Karsenty *et al*. have described as a validation as opposed to a design dialog [2]: the expert has the technical knowledge and the user has the domain knowledge. The expert proposes a solution that is accepted or modified by the user, who makes the final decision.

Cooperation is based on explanations. In a fully cooperative system, the e-coach modifies the user's perception of the problem, but there is also reciprocity [2]. The user can modify the e-coach by contributing his or her own explanations to the system. The approach that we are presenting in this paper focuses on the first aspect. The e-coach makes propositions (energy plans) and provides explanations. We have not yet considered the user feedback, which modifies the e-coach's perception of its environment.

In the next sections, we will first perform a quick review of various forms of explanations. We will then present our approach for generating explanations from the observed data in a smart home.

3.2. *The roles of explanations*

An explanation is a communication act between one or several persons. Its main objective is to directly increase the comprehension of the receiver. Beyond direct knowledge increase, an explanation may have several other objectives, as stipulated in [34] (a survey article that was the main inspiration for this section):

- To be able to predict similar events in the future [35].
- To be part of a diagnostic process [36]. Diagnosis is used to repair a malfunctioning system. It can also be used to reinforce an efficient or inefficient action to solve a particular problem. This association may help with choosing, or not choosing, the same action(s) to solve the same problem in the future.
- To justify an action.

All of these objectives, except the "diagnostic process for repair", are relevant to our goal. Helping users to predict the future and emphasizing the efficiency of their actions will help them to act correctly in similar situations.

3.3. *Different forms of explanation*

Historically, explanations, as a scientific domain, were primarily studied in social sciences (behavior explanations) and in philosophy (scientific explanations) [37]. For example, Aristotle identified two types of scientific knowledge: "knowledge about what" and "knowledge about why". "Knowledge about what" is descriptive whereas "knowledge about why" is explanatory. "It is one thing to know that each planet periodically reverses the direction of its motion with respect to the background of fixed stars; it is quite a different matter to know why" [38].

Aristotle considered that a scientific explanation should be a list of deductive arguments. This early influential vision has led to the widely used Deductive-Nomological model (D-N) [39]. A D-N model of explanation is similar to a logical proof. It involves three sorts of statement: statements about the initial conditions, statements about laws and theories (both called the explanans), and the observed statements that describe the phenomena to be explained (the explanandum). The explanandum is validly explained if it can be validly deduced from the explanans and the explanation is the deduction. In this process, all of the statements must be true: the explanandum (which is obviously true because it is observed), the initial conditions and the laws. For instance, “all gases expand when heated under constant pressure” is a law. An initial condition can be “a bladder balloon is heated” and the observation is “the size of the bladder balloon is increasing”. The observation can be deduced from the initial condition and the law is the explanation. One of the limitations of this approach is the definition of the “law”. A law should be a true generalization that does not suffer exception. Typically, a probabilistic phenomenon is not qualified as a law. The singular causal explanation of Michael Scriven [40] is also an example of a statement that will not meet the law criterion because it is not a generalization: “the impact of my knee on the desk caused the inkwell to tip over”.

New models of explanations have then been proposed to address the limitations of the D-N model. In particular, the logic formalism has been relaxed toward a variety of forms including textual representations, diagrams, and simulations. These forms may not capture the full complexity of a phenomenon but they serve as useful approximations of a phenomenon. Here, causality plays a central role.

Causality is usually present in everyday life explanations. Philosophers of science have used this notion to illustrate why some things explain others and not the opposite. Causal relationships can be deterministic (like two pieces of a mechanism) or probabilistic (as in economy, psychology, etc.). However, explanations do not rely solely on causal relationships. For instance, one can explain how a mathematical result is achieved, or why China is bordered by 14 different countries [34]. Nevertheless, when both causal and non-causal elements are present in an explanation, the causal elements dominate the patterns of judgment [41].

There are at least four kinds of causal explanations:

- Common cause. The same cause has several effects, like a cold may cause fever and a stuffy nose. It is often used in diagnosis.
- Common effect. Several causes converge to the same point. This is often the case in history. For instance, World War I was not caused only by the murder of Franz Ferdinand.
- Linear chain. There is a unique series of steps between an initial cause and a final effect. Linear chains are easy to understand but not very frequent in real life.
- Causal homeostasis. This is a causal relation that forms a stable cycle that reinforces itself.

Another way to differentiate between various types of explanations (D-N, textual, causal, a-causal, etc.) is the use of stances [42]. A stance is one way to set the frame for an explanation. A stance can be mechanical, design or intentional. The mechanical stance considers that the phenomenon to explain is composed of entities and is the result of their organization and interactions [43]. The explanation can be the mechanism itself (ontic) or the description of the mechanism (epistemic) [44]. The design stance considers that entities have functions and have been created on purpose. The explanation is based on that purpose and not on the mechanical functioning of the entity. Finally, in an intentional stance, the entity whose behavior is to be predicted is treated as a rational agent which, given its supposed beliefs and goals, will very likely act in order to reach these goals.

Our primary objective is to help users to better understand how they can modify their behavior and how they can improve their decisions with respect to their own comfort criteria. Due to the complexity of the building physics as well as to unconscious routines, users have difficulty in understanding what is happening and why they need to change their routines to improve their comfort, or how to make appropriate compromises between comfort and cost. Users do not need explanations as logical proofs (D-N), but as approximations that provide them with sufficient information about the current phenomena (epistemic mechanical stance). Explanations must not be generic but strictly related to their behavior as well as to the characteristics of their housing. As mentioned above, causality plays a central role. Generation of contextual and causal explanations is therefore at stake.

3.4. *From sensory data to causality*

End-users' housing is equipped with sensors that measure environmental variables such as temperature, humidity, CO_2 level, weather conditions, and the number of people in a room. Sensors also measure the user's actions such as turning on the heater or opening doors and windows. In our case, generating an explanation requires the analysis of the data flow provided by the sensors as well as the identification of the causal relationships between the actions of the users and variation of the environmental variables. In the next part of this section, we will considerer that a phenomenon is the value or the variation of the value of a variable. Some of the phenomena might be characterized as causes, some others as effects of these causes.

3.4.1. *Problem statement*

Causality from sensory data is difficult to model mathematically. Effects can be directly observed, but causal relationships cannot. Considering phenomena as events, a cause (*C*) always precedes the observation of an effect (*E*), but an effect (*E*) observed after (*C*) and correlated with it does not necessarily mean that (*C*) is the cause for (*E*). The "car allergic to vanilla ice cream" scenario illustrates this case [45]: a man used to buy ice cream after dinner for his family. He complained to General Motors that every time he bought vanilla ice cream, he had difficulty starting the car engine (other ice cream flavors were fine). General motors engineers finally found that the cause of the problem was vapor lock. Actually, it took less time to buy vanilla ice cream than for other flavors. As a result, the engine remained too hot for the vapor lock to dissipate. The co-occurrence of buying vanilla ice cream and the car not starting did not mean that buying vanilla ice cream was the cause of the car failure.

As illustrated in Figure 6, the co-occurrence (with a potential time delay *dt*) of two phenomena calls for several interpretations: precedence only (6a), direct causal relationship (6b), and consequences of a third phenomenon that may be outside of perception (6c). For instance, having the flu may first cause fever and then coughing. Ignoring the existence of viruses may lead one to believe that fever is the cause of the cough.

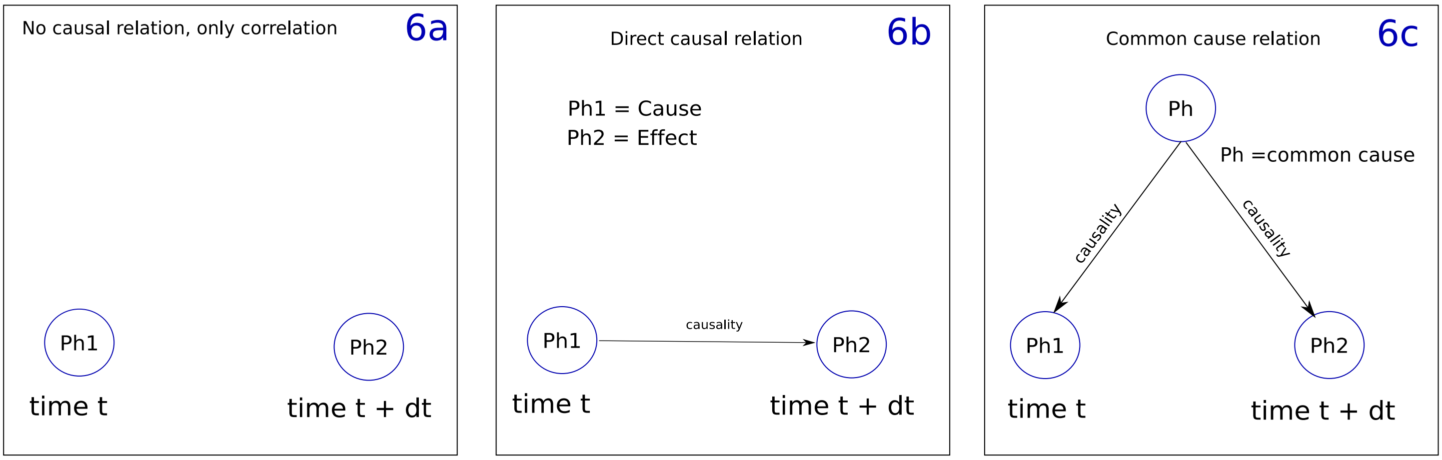


Figure 6. *Causal relationships between co-occurrent phenomena*

3.4.2. *Building a causal model from a bottom up approach?*

Extracting a causal model solely from time-correlated observations is a promising approach to the elaboration of a truly adaptive system. As almost no initial expert knowledge is necessary (everything is extracted from the data flow), such a system can be deployed in any housing and can learn a model of the environment from its interactions and observations. Constructivist or Developmental Artificial

Intelligence is an example of such approach. It was originally applied to robotics but there are some examples in the context of ambient intelligence (see [46] or [47] for instance).

However, learning a causal model from a purely bottom-up approach is a complex task. In particular, one cannot be sure that all of the necessary phenomena are observed in order to infer “common cause” causal relations. For a common causal relation (Figure 6c), is it necessary to find the common cause. If this common cause cannot be determined or observed, we might conclude the existence of a direct temporal relation between the two phenomena (as for Ph1 and Ph2 in Figure 6b). Direct relations are helpful for predictions. If Ph1 is observed, it is known that Ph2 will also be observed. But because one has a partial view of the world, one cannot predict when this relationship might fail. In addition, since the model is based on observations but not intrinsic to a mechanism, it cannot be used to determine which external intervention will modify its behavior. For instance, it can be observed that every morning, the rooster wakes up and sings, and soon after, the sun rises. There is of course no direct causal relationship between those phenomena but it can be used to predict the rising of the sun. Even so, there is no point in waking the rooster up earlier to get a longer day.

3.4.3. *Injecting knowledge in the building process*

Complexity rises from the size of the possibility space (all the possible causal models to explain the observed phenomena). One open question is how are our brains able to build complex representations with just sparse data? As stated by Tenenbaum *et al*. [48], “yet children routinely infer causal links from just a handful of events [49], far too small a sample to compute even a reliable correlation!”. As suggested by Plato, another source of information (as abstract background knowledge) must be available to help this inductive learning process. “Psychologists and linguists speak of constraints; machine learning and artificial intelligence researchers, inductive bias; statisticians, priors.” [48]. This abstract knowledge, or over-hypothesis, restricts the hypothesis or the model space at a less abstract level, reducing the complexity to find an appropriate model for explaining the data. Where does this over-hypothesis come from? Some authors suggest that over-hypothesis can be learned simultaneously with the model (the blessing of abstraction as opposed to the curse of dimensionality) using, for instance, hierarchical Bayesian approaches [48], [50], [51]. Hierarchical Bayesian approaches are relevant in the context of our work. One of the problems with these techniques is the quantity of data needed to obtain good results. Data is specific to every dwelling and cannot easily be mutualized. It is sparse and it may be necessary to have the system initially run for a long period of time in order to acquire enough data to get an initially usable causal model. As a result, the use of a knowledge model of the environment as high-level predefined expert knowledge has to be investigated.

In the context of our work, a knowledge model is a set of equations that describe a house, modeling the thermal properties of the walls, floors, and ceilings in terms of thermal resistance and thermal inertia, window surface area and orientation, etc. These models are able to predict the evolution of the physical variables (e.g. inside temperature, CO_2 concentration and humidity), by considering the value of these variables at the previous time step as well as the value of other environmental variables such as weather conditions and doors and windows opening. These thermal models rely on the analogy of electrical circuits. For instance, electrical resistance and capacitance represent thermal resistance and inertia. The electrical circuit is based on the characteristics of the house. The CO_2 model is a differential equation that uses room volumes, number of occupants, and CO_2 concentration in adjacent rooms. The various parameters of the thermal models are determined using genetic algorithms to fit the measurements of the sensors [52].

3.5. *Quality of an explanation*

Keil *et al*. [34] propose three dimensions to evaluate the quality of an explanation: circularity, relevance, and coherence. Others such as Kim *et al*. refer to the credibility of an explanation [53].

- A circular explanation is an explanation where the conclusion is used as part of the explanation. For instance, "this diet pill works because it helps people lose weight" (extracted from [34]). Complex circularities might be difficult to detect [54].

- Considering an explanation as a speech act, an explanation is relevant for a given goal if it has the appropriate level of details while not providing unnecessary or unrelated information. Particular care must be taken against ego-centrism as explanations may be based on what we believe the other person knows: the estimation of the other person's knowledge is often extrapolated from our own level of knowledge [55], which could be misleading, thus resulting in irrelevant explanations.

- Coherence means that the explanation is composed of a set of elements with some of them positively constraining others toward the effects.

The credibility of an explanation is related to the causal structure and is called "the causal diversity effect" [53]. A single cause might create several effects that in turn create several effects, etc., resulting in a causal tree. The causal diversity effect indicates that the further apart two final effects are in the tree, the more they are considered as a good justification of the initial cause. In 1847, the philosopher of science William Whewell called this "consilience" [56].

Another concept related to the evaluation of explanations, is the Illusion of Explanatory Depth (IOED). The IOED refers to the fact that people usually think, after having received an explanation about a system, that they understand it deeper than they really do [57]. Rose *et al*. have shown that the more people see or visualize a system component, the more they build mental simulations, and the more they believe they understand it.

Having presented the background knowledge related to persuasive technology and explanations, we now turn to the description of the Involved e-coach system, starting with the generation of explanations.

4. Generation of explanations in the Involved e-coach system

To illustrate our approach, we will use our laboratory office as a real-life experimental setting. This office is a single room with two outside windows, an inside window and an inside door connecting to an adjacent corridor. It is equipped with 27 sensors: door and window contact sensors; CO2 concentration; temperature within the office, in the adjacent corridor and outside; humidity; power consumption; etc. The office layout and the corresponding physical model are shown in Figure 7.

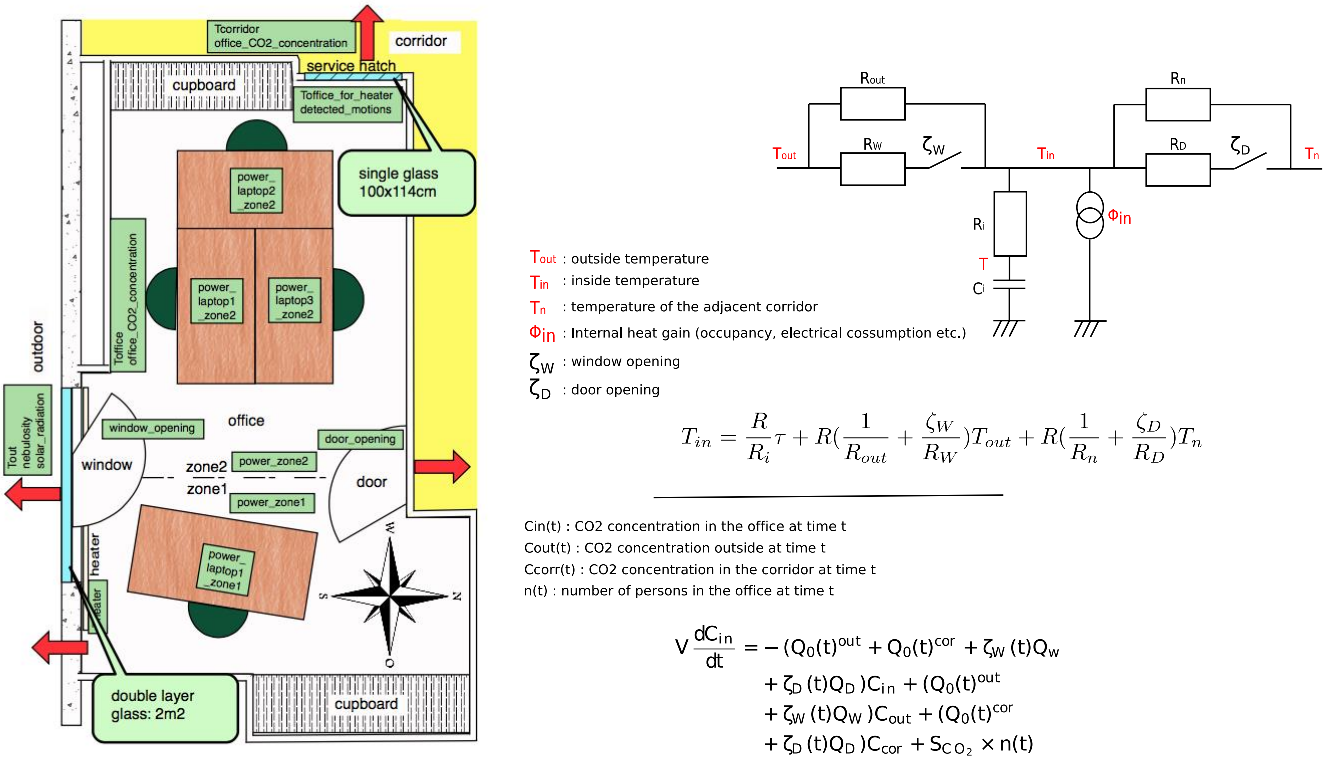


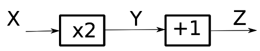
Figure 7. *Our experimental setting (laboratory office, left-hand side of the figure) with part of the corresponding physical model (right-hand side)*

The equations of the housing model are elaborated by experts whose goal is to predict the evolution of physical variables, not to model causality. These equations, which are not designed to produce explanations, do not represent causal relationships explicitly. As a result, causal relationships cannot be determined automatically from the analysis of the equations. The following example illustrates the difference between a causal model (represented as a diagram) and a mathematical equation (extracted from [58]). Two representations of a mechanism are considered: a set of two equations and a diagram that represents a chain of transformations (first a multiplier, then an adder) between three variables (*X*, *Y*, *Z*), going from X to Z:

Equation

$$\begin{cases} Y = 2 \times X \\ Z = Y + 1 \end{cases}$$

Diagram

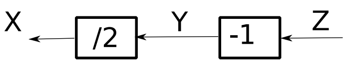


The equations and the diagram are not equivalent representations. The two equations above can be transformed into two **equivalent** mathematical equations that now correspond to the following new mechanism: a subtractor followed by a divider, going from *Z* to *X*.

Equation

$$\begin{cases} X = \frac{Y}{2} \\ Y = Z - 1 \end{cases}$$

Diagram



Finally, the first two equations and the second two equations can be summed up, providing two new mathematically **equivalent** equations, as shown below. These two new equations represent constraints on X, Y and Z but do not provide any indication as to how they influence each other.

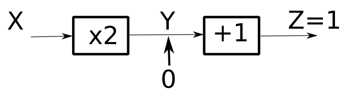
Equation

$$\begin{cases} 2 \times X - 2 \times Y + Z - 1 = 0 \\ 2 \times X + 2 \times Y - 3 \times Z + 3 = 0 \end{cases}$$

More precisely, consider an external intervention that imposes Y = 0. This post intervention can be modeled by adding a new constraint Y = 0, imposing Z = 1 (Eq 1). It can also be easily modeled in the second mechanism by imposing X = 0 (Eq 2). However, if Y = 0 is enforced in (Eq 3), nothing can be concluded for X and Z.

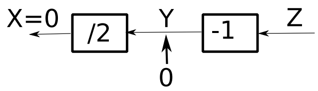
Initial mechanism | Post intervention | Diagram

$$\begin{cases} Y = \cancel{2 \times X} \\ Z = Y + 1 \end{cases} \qquad \begin{cases} Y = 0 \\ Z = Y + 1 \end{cases}$$



(Eq 1)

$$\begin{cases} X = \frac{Y}{2} \\ Y = \cancel{Z - 1} \end{cases} \qquad \begin{cases} X = \frac{Y}{2} \\ Y = 0 \end{cases}$$



(Eq 2)

$$\begin{cases} 2 \times X - 2 \times Y + Z - 1 = 0 \\ 2 \times X + 2 \times Y - 3 \times Z + 3 = 0 \end{cases}$$

We cannot say anything on X and Z (Eq 3)

The example above shows that causality cannot always be extracted from a formal analysis of the equations of the housing model. Instead, causality must be elicited by making virtual experiments on the mechanism described by these equations. Making a virtual experiment means defining a scenario, selecting variables' values (e.g. outside temperature) and using the physical equations to compute the missing values like air flow, heat flow, etc. The goal is to extract useful information from different measured and computed data and to represent them in a way that makes sense for the user.

In the next two sections, the virtual experiment approach is presented to generate direct differential explanations as well as differential explanations with influence.

4.1. *Direct differential explanations*

To reduce the risk of false causalities or circular explanations, the available variables' modeling phenomena have been organized into four groups (see Figure 8):

- Action: list of user's actions on the environment that have an impact on the comfort level (in our experimental setting, opening/closing doors and windows). Door or window opening is the percentage of time a door or a window has been opened during a one-hour time slot.

- Context: context can be representational or interactional [31], [33]. There is no absolute definition and it depends on the context of the problem. A representational approach has been selected because it is related to the current setting state. The e-coach context contains variables that have a role in the final decision while ensuring the stability property, which is not modified by the user energy related decisions. It is for instance the outside temperature T_{out}, the solar intensity, the temperatures in the neighbor zones (e.g. temperature of the corridor T_{cor}), the number of occupants, etc. A sequence of actions executed on two days but in a same context should have the same result. Opening the window on two similar summer days will have the same effect: heating the room. Opening the window on a summer or winter day will have opposite effects.

- Intermediate: these variables are either measured (e.g. indoor temperature T_{in}, CO_2 concentration) or estimated like the heat flow and air flow. These variables correspond to what physicists consider to be important in the causal chain.

- Effect: the conclusion of explanations that are directly related to human perceptions (indoor temperature and air quality, costs). These are the variables such as thermal dissatisfaction and air quality dissatisfaction experienced by the users.

Natural tendency of causal relations between the groups of variables are also pre-determined as expert abstract knowledge [48].

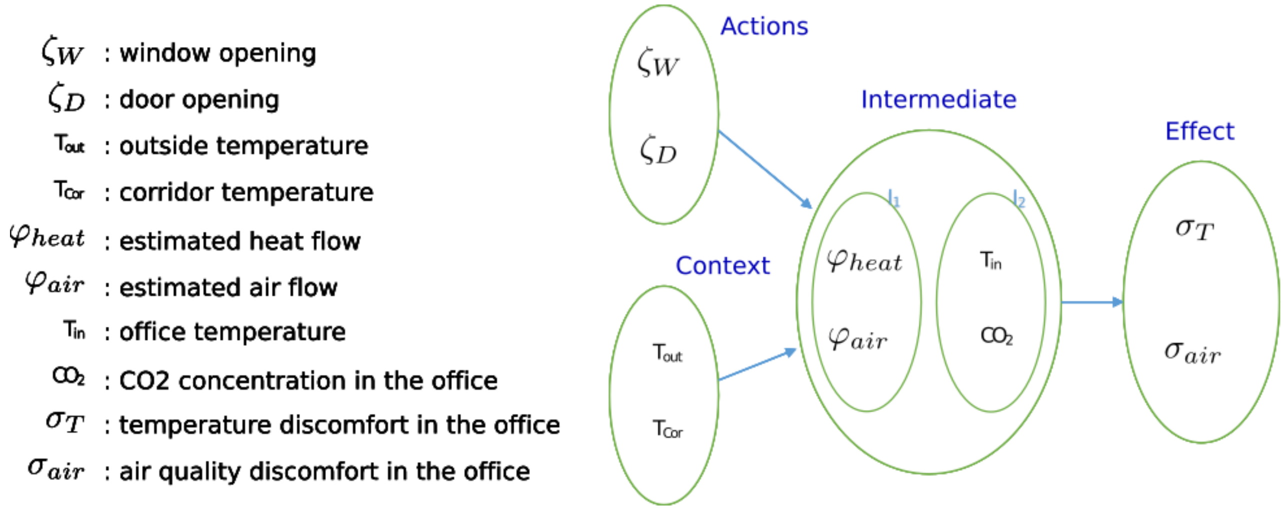


Figure 8. *Variable categories and main causal relations*

Users may not act "optimally" with respect to their own comfort criteria. In energy management, optimality is defined as "Pareto optimality" where any criterion (or objective) of a multi-criteria optimization problem cannot be minimized without increasing the others. In other words, compromises have to be made between conflicting criteria. In energy management, comfort levels and energy and environmental costs are typical conflicting criteria. For example, in winter, absence of heating would be optimal with regard to cost but not acceptable in terms of thermal comfort. As another example, good air quality may involve opening the window which in turn may lead to poor thermal comfort in winter and high energy cost. Our hypothesis is the following: by generating "differential explanations", end-users will be able to find the preferred optimal trade-off between conflicting criteria.

A differential explanation is constructed by analyzing the difference between two scenarios defined in a same context (scenarios said to be context-similar) for the sake of meaning: (1) a past scenario for instance can be compared with an alternative, possibly optimal scenario, or (2) a forecasted scenario compared to an alternative scenario. The main objective is to evaluate the difference in perceptions associated with the difference in actions between both scenarios (the difference in perceptions is the difference in effects associated with the difference in actions because both scenarios are in the same context). By carefully selecting the two scenarios that are compared, various explanations can be computed. The two following scenarios are firstly be considered: the "recorded history scenario" includes: (1) the set of actions performed in real life by the users during the day (doors and windows opening in our experimental setting), (2) contextual environmental information (weather, time, etc.), and (3) measured effects (air quality and temperature). The "Pareto optimal scenario" is selected from the Pareto front, i.e. from the set of optimal solutions computed using a differential approach for the following criteria: air quality, thermal comfort, and energy cost (see [58] for details).

A scenario is compared to a reference scenario (the two scenarios are what has been done and what would have been the best scenario) over a period ranging from 8 am to 8 pm (office hours) divided into 12 one-hour time slots. The doors and windows opening values are the percentage of opening time during each slot. The other values such as temperature or CO_2 are the average values during the same time slot. The difference between the two scenarios is the computation of three values for every one-hour time slot:

1. Difference in terms of actions: the difference of opening time percentage of doors and windows for each corresponding time slot of scenario 1 and scenario 2.

2. Difference in terms of effects: the difference in air quality and temperature comfort for each corresponding time slot of scenario 1 and scenario 2.

3. Difference in terms of intermediate effect variables. Some of the intermediate effect variables (e.g. air or heat flow) are not directly measured by sensors. Therefore, they are not available in the “recorded history scenario”. Instead, we draw on the physical model, which uses sensory data as input, to estimate the intermediate effect variables.

It is then necessary to transform the quantitative values into qualitative information. For instance, the sentence “shutting the door at 2 pm will cause an important decrease in the airflow and a significant decrease in the air quality level” is easier to understand than “a difference in airflow of 30% will lead to a difference in CO_2 concentration of 400 ppm (parts per million)”. The transformation from quantitative data to qualitative information is performed by dividing the value domain of a variable into seven sub-domains (three positive levels, three negative levels and one no-change level). An example of the differences between the “recorded history scenario” and the “optimal Pareto scenario” is illustrated in Figure 9.

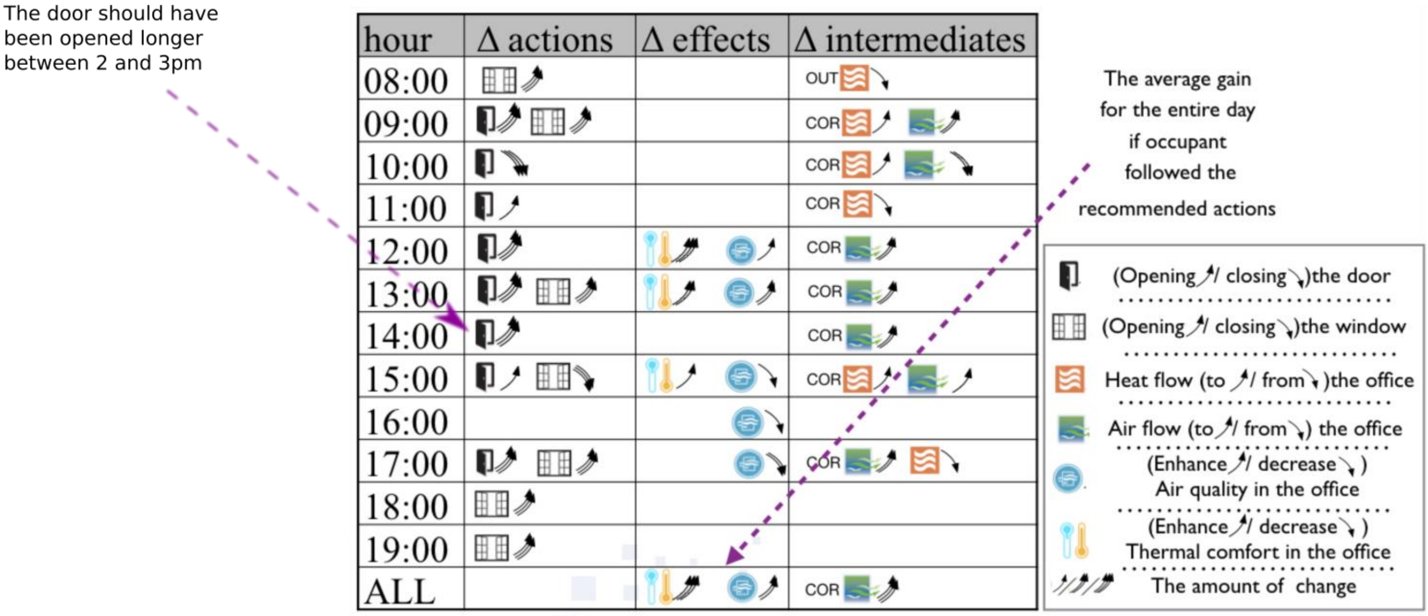


Figure 9. *Difference between what really happened during a day in the office and what could have happened if the scenario extracted from the Pareto front would have been followed. For instance, between 2 pm and 3 pm, the office door should have been opened longer, thus increasing the air flow from the corridor to the office. At the end of day, the air quality and temperature comfort would have been increased*

The explanation that can be extracted from the difference between the history-based and the Pareto-optimal scenarios is what Aristotle called “knowledge about what”. It is descriptive, not causal. For instance, from 8 am to 12 am, the system points out that the windows and doors should have been opened for a longer or shorter period of time. At 12 am, the temperature and air quality would have been increased but there is no explicit link between these two. This explanation informs the users about how they should change their behavior along with the gain that they can expect from this change. On the other hand, the explanation does not provide any indication as to which action is really responsible for a given effect. For instance, is it really important to close the door between the office and the corridor between 10 am and 11 am? If the end-user is out of the office at this time and if he or she does not intend to come back to close the door, will it ruin the air and comfort qualities later in the day? To answer these questions, we need to go deeper into the explanation process and compute the impact of each action and its resulting effects. This problem is addressed with “differential explanations with influence”.

4.2. *Differential explanations with influence*

The actions proposed by the system in the previous section do not have the same importance in terms of impact. Some of them might be skipped if necessary, and some of them should be performed because of their strong influence on a particular criterion. To evaluate the influence of an action A, we compute the difference between the Pareto-optimal scenario and the Pareto-optimal scenario where A is replaced by the action that has been performed effectively by the user. Both scenarios are simulated using the physical model of the house. The difference between the effects indicates the influence of not performing the recommended action A. The result of this influence computation is displayed in Figure 10.

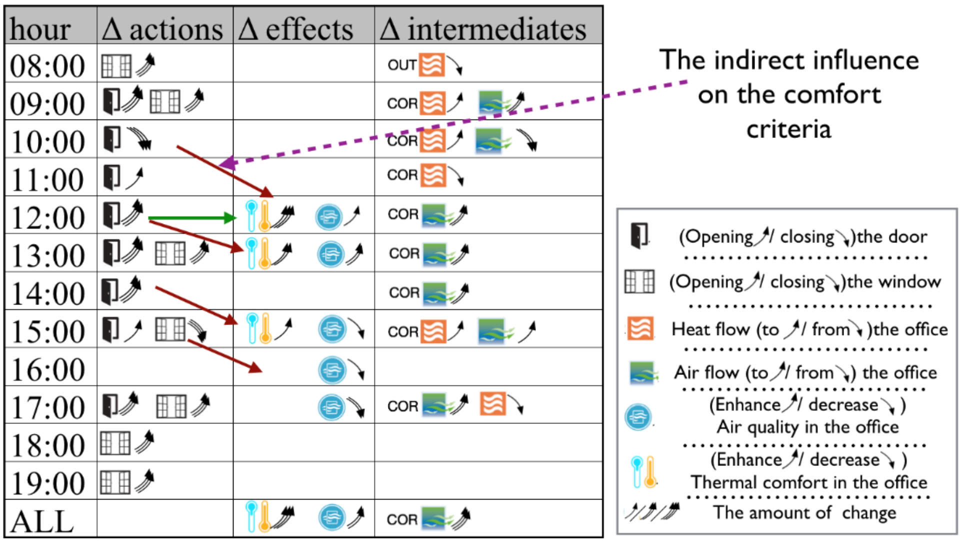


Figure 10. *Differential explanations with influence between actions and their final effects. Opening the door longer at 10:00 will increase air quality and temperature comfort at 12:00 for instance*

Using the equations of the knowledge model, it is possible to generate cause–effect relations between actions and final effects, but the causality between the different levels of the intermediate variables and the final effects are unreachable because their changes cannot be monitored with a knowledge model (see Figure 11). Knowledge models are based on physics (see [52] for instance).

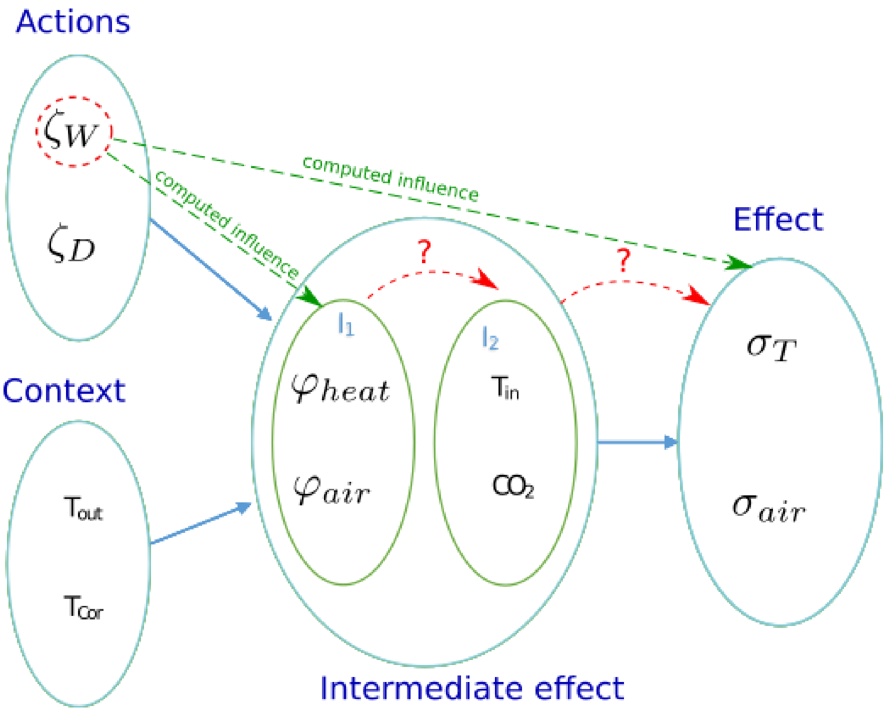


Figure 11. *Missing knowledge. The physical equations cannot be used to simulate the effects of intermediate variables*

Integrating relations between the intermediate and the final variables is important to provide users with complete explanations. This can be achieved by injecting expert knowledge in terms of rules that represent potential causalities as well as impossible ones. For instance, heat flow may have an influence on air temperature but not on CO2 concentration. This can easily be done because the expert has a very good knowledge about the nature of those variables.

By integrating calculated influence and potential influence, a full diagram for the whole system can be automatically elaborated. Part of this diagram is represented in Figure 12.

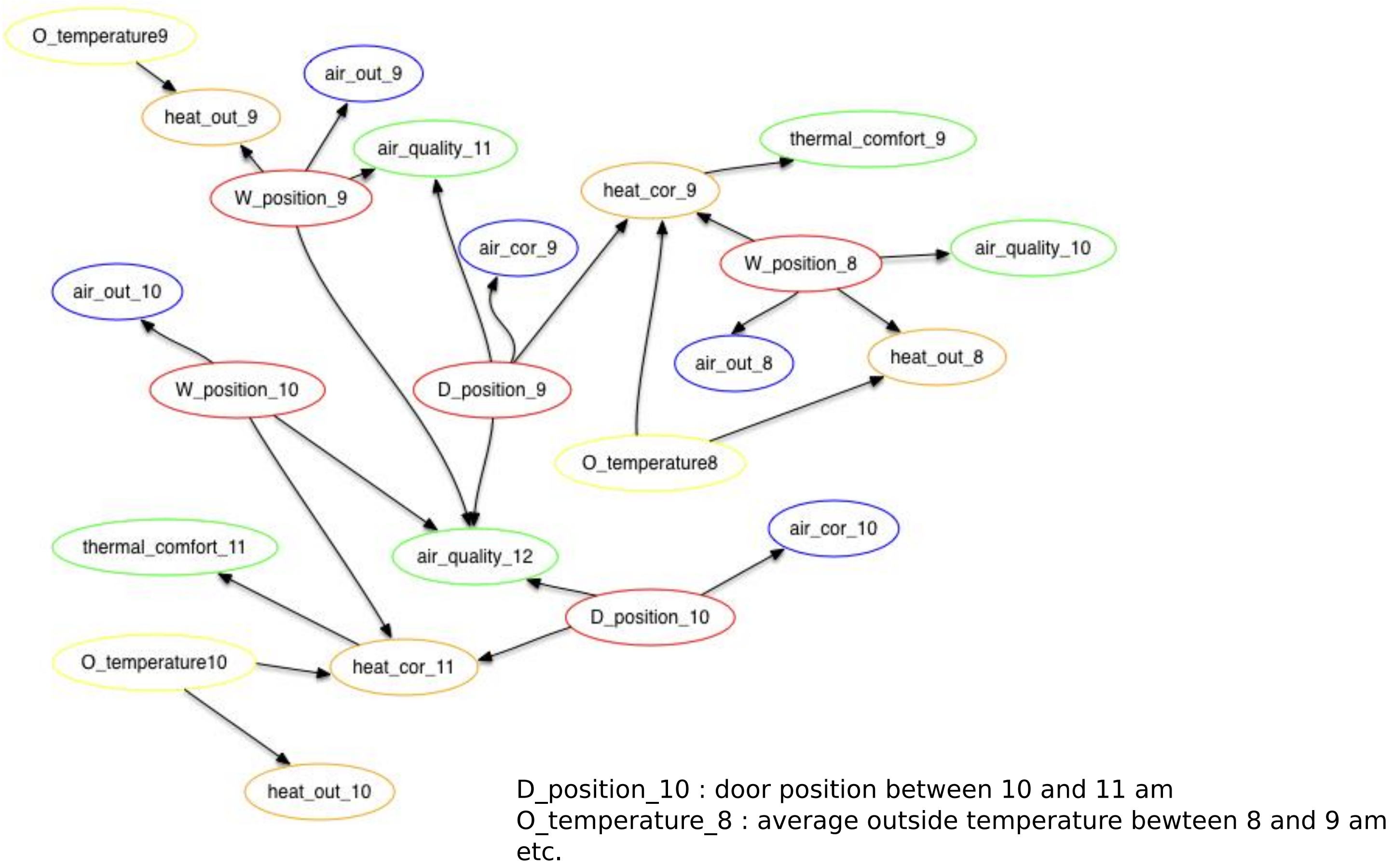


Figure 12. *Part of the path diagram (relations between all variables) covering one full day*

4.3. *Natural language text generation*

Reading complex tables and graphs, such as those presented in Figures 10 and 12, may be a rather abstruse experience for the user. Users may prefer statements expressed in natural language, provided that these statements are well written and smart enough, for example not using the same wording repeatedly.

Natural Language Generation (NLG) is a Natural Language Processing (NLP) sub-field concerned with the generation of texts from non-linguistic data [60]. Typical systems take a high-level representation of input data, select and structure the final information to convey (macro-planning), make a lexical and syntactic transformation to this structured information (micro-planning), and finally linearize it as text (surface generation). This classical pipeline has mainly been approached by expert and grammatical rules [61], statistical models [62], or machine learning [63] in areas as diverse as complex medical data summarization [61], prose generation [62], or image captioning [63].

A NLG system is both a dialog system and a Machine Translation (MT) system. Most recent stochastic MT systems perform NLG through shallow language models, but there are also MT systems that propose a deep general-purpose NLG module using a high-level internal representation. This high-level internal representation is extracted from a linguistic analysis of the target language analysis and is used to generate a text that complies with its grammatical rules.

We have chosen the Ariane-Heloise MT system, a reengineered version of the widely known Ariane MT system [64]. Ariane-Heloise links together several transducers that transform a source text into a decorated tree, then transform this tree into a linguistic tree structure, and generate a target text from this structure. During the process, various linguistic operations are carried out: morphological and structural analysis, lexical transfer, syntactic and morphological generation. The decoration of each node of a tree is a combination of values for a set of declared variables and allow for coding linguistic properties and relations. The following is an example of a generated textual explanation (in French): “Dans le créneau horaire 12h–13h, si vous aviez laissé la porte ouverte beaucoup plus longtemps, le confort thermique aurait augmenté beaucoup, la qualité de l'air un petit peu et le créneau 13h–14h aurait été impacté, il y aurait eu un courant d'air sensible vers le couloir”. More details can be found in [65].

5. Interaction with the Involved e-coach system

Given the explanation engine presented above, a user interface is mandatory to convey the explanations and to support the user in the energy management process. However, information display and interactivity are not enough to involve the user in such a process. If the user is unaware, not motivated, or unable to control her/his energy consumption, such a system is useless. Therefore, a user interface has been designed for Involved that supports awareness, aiming to persuade and help users to change their behavior in order to reduce their energy consumption while feeling comfortable. In the following, we focus on such a user interface and illustrate our design principles.

5.1. *User interface requirements*

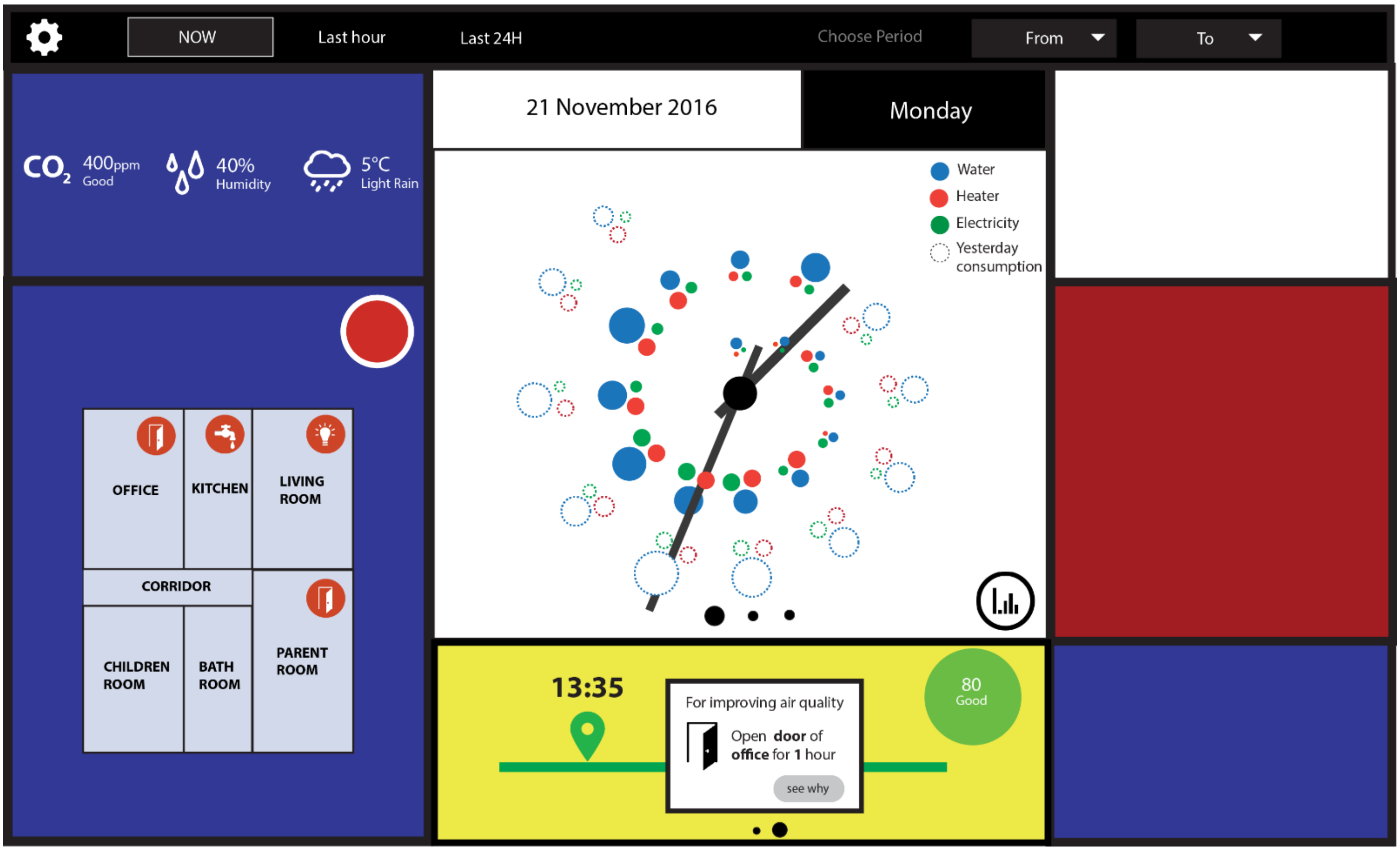


Figure 13. *Design concept inspired by Mondrian’s style*

The user interface of Involved is intended to be deployed on a tablet installed on a wall of the living room or in the entrance hall. Thus, several issues have to be considered: (i) standing in front of the tablet may not be comfortable compared to sitting at a desk; (ii) depending on the period of the day, the user may not be available to interact with the system (e.g. leaving home in a hurry to go to work); (iii) over time, users may not pay attention to the device any more.

To address these issues, the user interface (UI) must satisfy the following two design requirements: (R1) the UI must be eye-catching; (R2) multiple levels of interaction for different contexts of use (e.g. key moments of the day) must be supported. According to FBM (cf. section 2), these two requirements aim to maintain high motivation and ability. For example, by embedding utility functions such as time and weather forecast, the UI should catch the eye at key moments of the day (e.g. looking at the clock before leaving home). Consequently, it should maximize (i.e. make the user aware of) the persuasive features of the systems such as reminders of actions to reduce energy consumption as well as greetings and rewards in order to maintain appropriate behaviors.

5.1.1. *Designed for daily life: catching the eye at the right time (R1)*

One compelling approach to potentially impact the motivation of individuals is to offer an interactive system for daily life. Our suggestion is to bring daily activities and objects such as checking the time, checking the weather, or decorative elements into the UI. Figure 13 illustrates the design concept in the style of Mondrian. To satisfy requirement R1, it is hypothesized that an artistic user interface (e.g. a painting) that embeds daily utility services (e.g. a clock) would catch the eye at any moment of the day and be attractive. The location of the device is also important. In their long-term field study of the Digital Family Portrait[1], Rowan and Mynatt [66] installed the device at the entrance of the participant's kitchen. They observed that the choice of the location was key for it to be readily visible and therefore effective.

5.1.2. *Three levels of user interaction for three contexts of use (R2)*

To meet requirement R2, the user interface has been designed to support different contexts of use. Here, "context of use" is defined as "time or period of the day, and user's expertise and interest in energy management". Depending on the time of day, users may not have enough time to use the system. Therefore, the UI must be designed to support very brief user interactions as well as longer periods of interaction. We hypothesize that such an adaptation to the context of use is a means, according to the FBM, to increasing ability as well as motivation by providing customized messages such as explanations. These functionalities would also support maintenance. Three classes of UI properties have been selected for three classes of use:

Glanceable or "no-click" UI. A glanceable display is designed to provide quick and easy feedback to the user [Gouv16]. Going one step further, we envision our user interface to provide the user with synthetic and relevant information depending on the time of day (e.g. checking before leaving home). For instance, for the Digital Family Portrait [66], the key moments were leaving for and returning from work. At a glance, the appearance and size of the butterfly icons informed the participants about the relative's current level of activity.

One-click UI. The user achieves a very short user interaction with the system (e.g. requesting details, validating a suggestion). In terms of persuasion, this level of interaction constitutes "baby steps" to involve the user in the process. For instance, the e-coach engine may require the user to indicate an appropriate compromise between comfort level and energy cost.

Zoomable UI. The user performs a longer interaction to explore the state of the home or the capabilities of the management system. It is assumed that it is achieved for a suitable (e.g. comfortable) context of use (e.g. sitting at a desk). To address this situation, the overall design of the UI is based on the principles of zoomable user interfaces [67]. A zoomable user interface offers users the capability to continuously change the scale and level of detail of some information without losing the informational context.

With these three types of UI properties (glanceable, one-click, and zoomable UI), it has been hypothesized that we can target a wide range of users with different levels of implication or different interests in energy management as well as different levels of expertise.

---

[1] The Digital Family Portrait is a picture of users' elderly relative who lives remotely, framed with butterfly icons that indicate the level of activity of the relative during the day.

5.1.3. *User interface in detail*

As shown in Figure 13, Mondrian's style has been adopted as the baseline for the aesthetics of the UI. Similar to Microsoft Windows 8, one property of this style is to split the screen space into tiles. Here, the tiles represent a block of information that corresponds to a specific task. At the center of the interface, a 24-hour clock provides an overview about the energy consumption of the last 24 hours.

At a glance. A zoomed-out tile provides synthetic information that requires no or very little user interaction. If more detail is needed, using semantic zoom interaction techniques, a tile is expanded to provide more space at the expense of the other tiles. Three major tiles are identified: a spatial view of the habitat (blue tile on the left-hand side), a temporal view of the energy consumption (clock-based white tile), and the e-coach view (yellow tile). The white and yellow tiles both support sliding gestures to present complementary views. On the left-hand side, the blocks filled in blue provide information about the outside conditions (top-left blue rectangle: air quality, humidity percentage, weather conditions) and about the interior of the dwelling. The latter is represented through a map of the home where a rectangle represents a room or a corridor. As detailed in the following, red and green circles indicate the global status of the home as well as the status of each room.

At the center of the screen, the white tile displays a clock (a utility service) associated with time-based synthetic visualizations. The spiral-based visualization presented here is only one possible illustration. Additional complementary visualizations can be obtained using a sliding gesture performed at the bottom of the tile. These visualizations represent information about energy consumption, domestic activities, daily objectives, etc.

The yellow tile (bottom center) is dedicated to interaction with the e-coach. For instance, a synthetic view may be a representation of the suggested next actions to be achieved within the current hour slot of the day. With a click, the user can obtain detailed explanations about the action and its consequences.

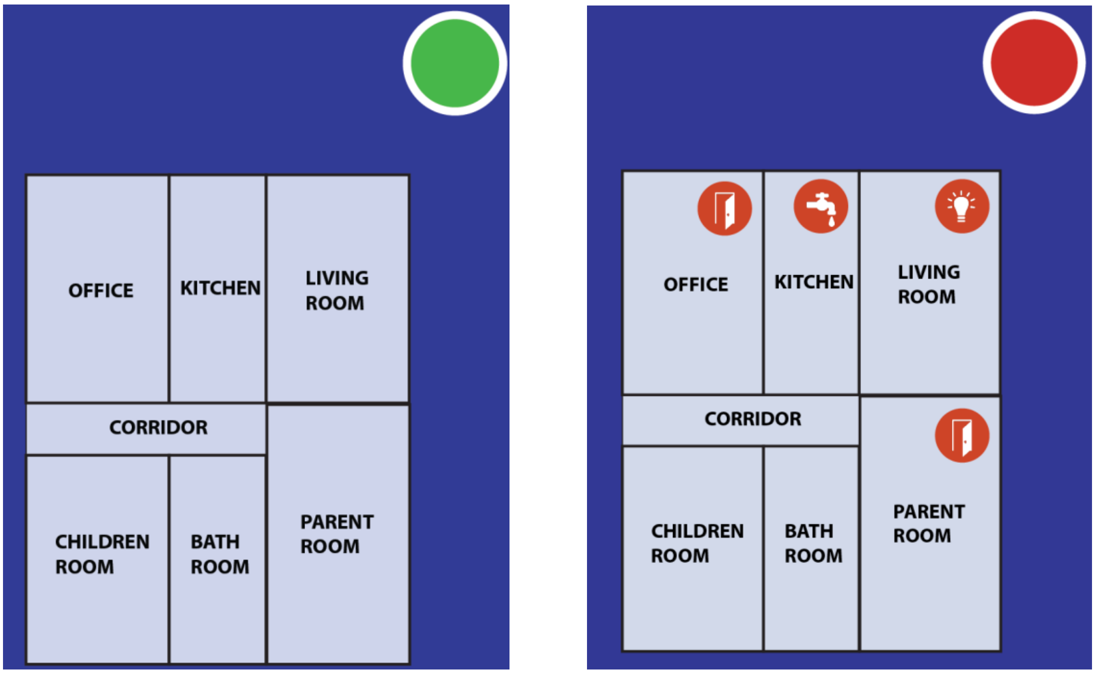


Figure 14. *a) No issues: green light (left); b) issues highlighted with a red icon (right)*

Use of everyday metaphors. When people face something new and want to understand it, they usually try to compare it to things that they already know in order to fit it into their knowledge structure. Thus, the use of metaphors from everyday life in a UI may facilitate learning. The metaphors can be based on an activity or an object that is familiar to the user. In our UI, we have used the traffic light metaphor to promote a glanceable (i.e. concise) user interface: “at a glance, I can check if everything is okay in my home”.

As explained above, this part of the user interface (Figure 15) represents a logical representation of the habitat (less accurate than a real two-dimensional map). The traffic-light metaphor is used as follows: if the system detects an undesired event (i.e. lights are on in a room with no presence), a red circle (Figure 14b) is drawn in the relavent room associated with an icon (e.g. a light bulb means that lights are on; an open door means that, according to the e-coach, a door should stay shut in order to optimize temperature in the room and/or maintain air quality). The circle located in the top-right corner provides a global overview. Therefore, a green circle means that everything is okay (Figure 14a).

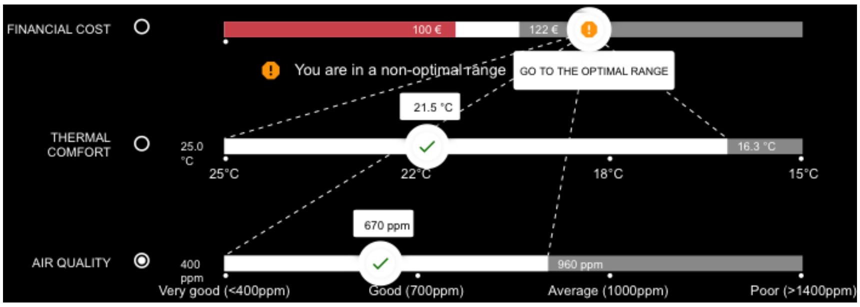


Figure 15. *Finding the appropriate compromise between conflicting criteria using the Trade-Off-Pareto sliders (https://pareto-sliders.firebaseapp.com/)*

Designed for everyday tasks. Another method to keep users motivated is to give them full and easy control of their dwelling, in particular by simplifying the complex notion of multi-criteria optimization. Our solution is to hide this complexity in the form of an everyday activity such as navigating on a map or manipulating familiar widgets.

Figure 15 shows the Trade-Off Pareto sliders (TOP-sliders for short) based on the composition of familiar widgets (here, sliders), extended with the appropriate features to allow users to find the preferred trade-off between thermal comfort, cost, and energy waste. A preliminary semi-controlled empirical study with 16 participants using three interdependent criteria shows that the TOP-sliders can serve as a sound basis to support non-expert users in making informed compromises [68]. We can make an analogy here with what Brézillon and Pomerol call the asymmetry in an Intelligent Assistant System [31]: the system provides support by proposing satisfying solutions (the Pareto front) and arguing for them. The user makes the final decision.

The TOP-Sliders are composed of a set of parallel sliders, each of which is associated with a criterion of the optimization problem. As with standard sliders, the position of the cursor on a slider denotes the currently selected value for that criterion. The sliders are augmented with tightly coupled visual features to support decision-making through a “what if” process. These features represent (1) the interdependence between the criteria, (2) the impact of the modification of the value of one criterion on the others, and (3) for each criterion, the current ranges within which its values may fall, each range resulting from the Pareto front calculated for the current optimization problem.

Color coding is used to discriminate the three sorts of “Pareto ranges”: white denotes a range of values situated on the Pareto front; grey, a range of feasible but non-optimal values; and red for ranges of unfeasible values. For example, in Figure 15, the choice for thermal comfort (21.5°C) and air quality (670 ppm) is optimal whereas financial cost (greater than 122€) is not. The interdependence between the criteria is made explicit with pairs of dashed white lines where a pair pops up when a cursor is selected, and links this cursor to the boundaries of the optimal range of the other sliders. For example, in Figure 15, the cursor of the financial cost is currently selected: two pairs of lines have appeared to show the impact of the current choice on the range of the optimal values for the other two criteria. In addition, as the cursor of a slider is moved, the ranges of the other criteria are updated according to the underlying Pareto model.

Tight coupling between cursor movements and Pareto ranges makes explicit the impact of the selection of one criterion value on the remaining criteria. In order to allow users to explore trade-offs that are not necessarily Pareto-optimal, moving one cursor does not move the other cursors. To improve affordance and intelligibility for situations where the cursor for a criterion lies outside of the optimal range, an additional textual label coupled with the appropriate icon is displayed below the appropriate interval to explain the situation. In addition, a corrective action is proposed as a speed-up button (see Figure 15 for the Financial Cost slider). When clicked, the cursor is automatically moved to the nearest optimal value.

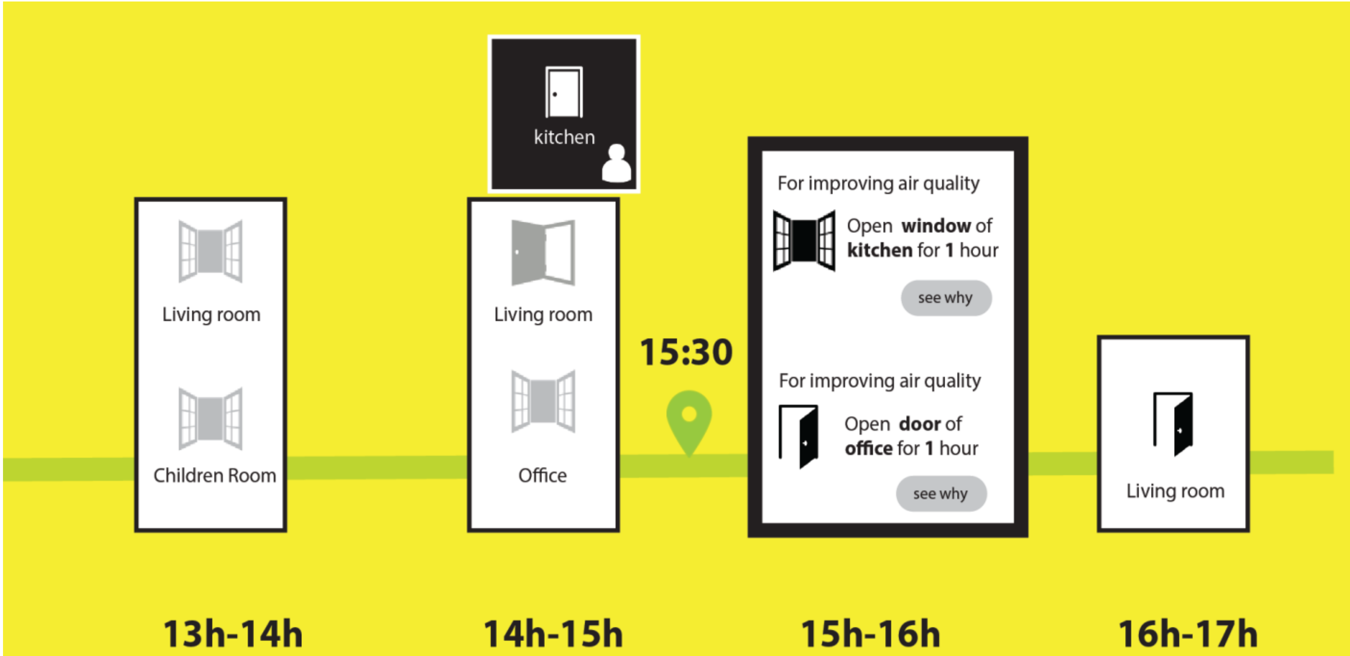


Figure 16. *Action plan recommended by the e-coach*

Once the trade-off between the criteria has been defined, the e-coach engine generates a list of recommended actions. Figure 16 shows the action plan recommended for the day. As explained in section 4, it represents the optimal scenario computed by the explanation engine. Clicking on the "see why" button attached to a particular action gives access to explanations about the action and its consequences[2]. Black boxes denote user's actions that are not part of the optimal scenario (e.g. closing the kitchen door between 2 pm and 3 pm). The interaction metaphor is based on that of well-known Geographical Information Systems (GIS), typically Google Maps: the action plan corresponds to the route recommended by the GPS with landmarks on the way as well as the current location, which in our case is a temporal location. As shown in Figure 16, it is currently 3:30 pm, past actions are greyed out, and the next actions to be performed between 3 and 5 pm are zoomed out.

[2] At the time of writing this, the "see why" button is not connected to the natural language text generator described in Section 4.3.

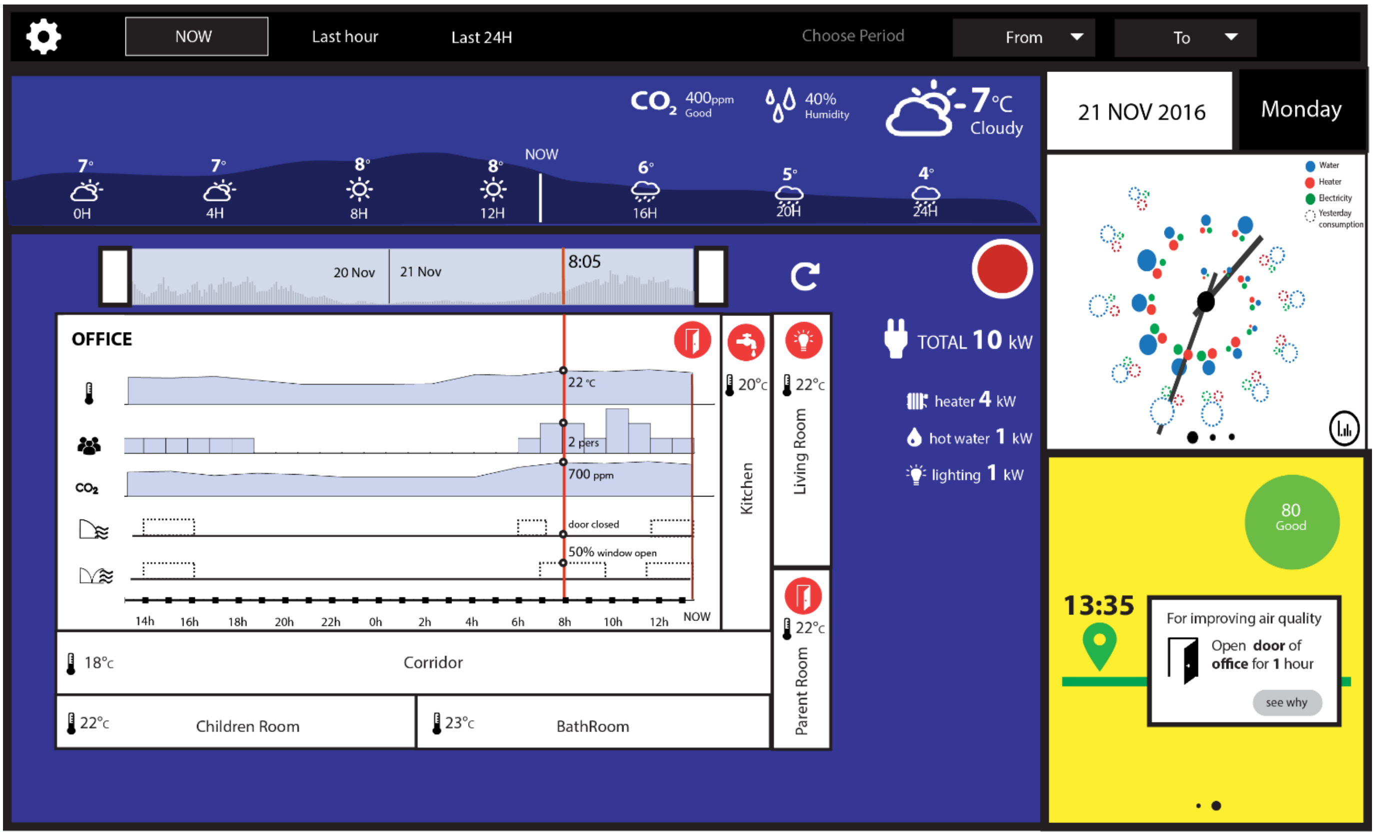


Figure 17. *Zoomable user interface: the focus is currently set on the office room while contextual temporal information and action plan information are still observable*

Multiple levels of interaction: zoomable UI. Hornbæk *et al*. [67] define zoomable user interfaces as interfaces that "*organize information in space and scale, and use panning and zooming as their main interaction techniques*". This interaction technique has been applied to support the three types of user interface properties and several levels of detail. Scaling is applied on tiles: double-clicking (or tapping) on a tile expands its space; the other tiles are shrunk and organized to keep the "informational context" visible. Figure 17 shows an expansion of the left-hand blue tile of Figure 13. The map of the home is augmented with graphs that show the temporal evolution of relevant indicators (e.g. indoor temperature, level of CO_2).

6. Conclusion

The conceptual contribution of this article is threefold: (1) generation of explanations for the user, including differential explanations; (2) design principles of persuasive user interfaces (UI) for energy management systems; (3) mapping of persuasive UI principles into concrete interaction techniques and metaphors.

From the technical perspective, an early prototype of the Involved e-coach system has been developed that provides end-users with a 24-hour plan of recommended actions along with contextual explanations that justify each action. This plan satisfies the user's preferred compromise between thermal comfort, air quality, and financial cost specified by the user. Users can edit the plan (e.g. suppress an action), skip some actions, perform additional actions, or even change their preferred compromise, and be informed in real time of the consequences on energy consumption and comfort.

In practice, in its current version, the generation of explanations has been tested while deployed in a controlled environment (that of an office), and the user-centered evaluation of the UI elements has also been performed in a controlled setting (again, that of an office) where the behavior of the explanations generator was simulated. Whereas these early steps are necessary to detect the basic technical flaws and limitations of a system, they must be completed by longitudinal experiments performed in real-world settings. Based on previous research results and field studies on home automation, it is expected that users will be interested in the system recommendations in the very first months of use, provided that they are useful, robust, and trustworthy. However, users can play with the TOP-sliders to express their preferred compromise and explore the solution space provided that the required attentional cost [69] is not too high. In the long run, either the e-coach is not used anymore, or the settings (compromises) are good enough and the system will run in the background as calm and optimal technology until some exception occurs. In this case, the Involved system could be called upon to specify a new compromise relevant to the current situation, but there is no guarantee that the user will be motivated enough to address the exception. Clearly, these hypotheses on future use will need to be validated experimentally.

Beyond those contributions and their coming evaluation, we also argue for integrated research combining expertise from different fields (e.g. physical buildings) as well as for openness toward artistic design, for example.

Acknowledgements

This work benefits from the support of the INVOLVED ANR-14-CE22-0020-01 project of the French National Research Agency. It is also supported by the French National Research Agency in the framework of the the EquipEx program AmiQual4Home ANR -11-EQPX-00 and the "Investissements d'avenir" program (ANR-15-IDEX-02) as the cross disciplinary program Eco-SESA. This work is also partially supported by the project (DST-INRIA/2015-02/BiDEE/0978) of the Indo-French Centre for the Promotion of Advanced Research (CEFIPRA|IFCPAR).

Bibliography

[1] "Chiffres clés de l'énergie - Edition 2016," 2016. [Online]. Available: www.statistiques.developpement-durable.gouv.fr/fileadmin/user_upload/Datalab-13-CC-de_l-energie-edition-2016-fevrier2017.pdf.

[2] L. Karsenty and P. J. Brézillon, "Cooperative problem solving and explanation," *Expert Systems with Applications*, vol. 8, no. 4, pp. 445–462, Jan. 1995.

[3] B. Fogg, "Persuasive Computers: Perspectives and Research Directions," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 1998, pp. 225–232.

[4] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179–211, Dec. 1991.

[5] S. U. Press, "A Theory of Cognitive Dissonance | Leon Festinger." [Online]. Available: http://www.sup.org/books/title/?id=3850.

[6] K. Torning and H. Oinas-Kukkonen, "Persuasive System Design: State of the Art and Future Directions," in *Proceedings of the 4th International Conference on Persuasive Technology*, New York, NY, USA, 2009, pp. 30:1–30:8.

[7] F. D. Davis, "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989.

[8] J. Macinnis, C. Moorman, B. J. Jaworski, D. J. Maclnnis, C. Moorman, and B. J. Jaworski, "Enhancing and Measuring Consumers' Motivation, Opportunity, and Ability to Process Brand Information from Ads," *Journal of Marketing*, pp. 32–53, 1991.

[9] B. Fogg, "A Behavior Model for Persuasive Design," in *Proceedings of the 4th International Conference on Persuasive Technology*, New York, NY, USA, 2009, pp. 40:1–40:7.

[10] J. O. Prochaska, C. C. DiClemente, and J. C. Norcross, "In search of how people change. Applications to addictive behaviors," *Am Psychol*, vol. 47, no. 9, pp. 1102–1114, Sep. 1992.

[11] J. O. Prochaska and C. C. DiClemente, *The transtheoretical approach: crossing traditional boundaries of therapy*. Homewood, Ill.: Dow Jones-Irwin, 1984.

[12] B. J. Fogg, *Persuasive Technology: Using Computers to Change What We Think and Do*. Elsevier, 2003.

[13] E. Arroyo, L. Bonanni, and T. Selker, “Waterbot: Exploring Feedback and Persuasive Techniques at the Sink,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2005, pp. 631–639.

[14] A. Gustafsson and M. Gyllenswärd, “The Power-aware Cord: Energy Awareness Through Ambient Information Display,” in *CHI ’05 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2005, pp. 1423–1426.

[15] P. Petkov, S. Goswami, F. Köbler, and H. Krcmar, “Personalised Eco-feedback As a Design Technique for Motivating Energy Saving Behaviour at Home,” in *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, New York, NY, USA, 2012, pp. 587–596.

[16] J. Kjeldskov, M. B. Skov, J. Paay, D. Lund, T. Madsen, and M. Nielsen, “Eco-Forecasting for Domestic Electricity Use,” in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA, 2015, pp. 1985–1988.

[17] C. Elsmore, M. Wilson, M. Jones, and P. Eslambolchilar, “Neighbourhood Wattch -Community Based Energy Visualisation For The Home,” in *Nundge & Influence Through Mobile Devices workshop (NIMD)*, 2010.

[18] J. Froehlich *et al.*, “UbiGreen: Investigating a Mobile Tool for Tracking and Supporting Green Transportation Habits,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2009, pp. 1043–1052.

[19] L. E. Holmquist and T. Skog, “Informative Art: Information Visualization in Everyday Environments,” in *Proceedings of the 1st International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia*, New York, NY, USA, 2003, pp. 229–235.

[20] C. Fan, J. Forlizzi, and A. K. Dey, “A Spark of Activity: Exploring Informative Art As Visualization for Physical Activity,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, New York, NY, USA, 2012, pp. 81–84.

[21] T. Nakajima and V. Lehdonvirta, “Designing motivation using persuasive ambient mirrors,” *Pers Ubiquit Comput*, vol. 17, no. 1, pp. 107–126, Jan. 2013.

[22] G. Cockton, “A Development Framework for Value-centred Design,” in *CHI ’05 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA, 2005, pp. 1292–1295.

[23] R. M. Ryan and E. L. Deci, “Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being,” *Am Psychol*, vol. 55, no. 1, pp. 68–78, Jan. 2000.

[24] E. Costanza, S. D. Ramchurn, and N. R. Jennings, “Understanding domestic energy consumption through interactive visualisation: a field study,” in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, New York, NY, USA, 2012, pp. 216–225.

[25] S. Gregor and I. Benbasat, “Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice,” *MIS Quarterly*, vol. 23, no. 4, pp. 497–530, 1999.

[26] M.-C. ROUSSET and B. SAFAR, “Negative and Positive Explanations in Expert Systems,” *Applied Artificial Intelligence*, vol. 1, no. 1, pp. 25–38, Jan. 1987.

[27] C. L. Paris, “Generation and Explanation: Building an Explanation Facility for the Explainable Expert Systems Framework,” in *Natural Language Generation in Artificial Intelligence and Computational Linguistics*, C. L. Paris, W. R. Swartout, and W. C. Mann, Eds. Boston, MA: Springer US, 1991, pp. 49–82.

[28] B. M. Muir and N. Moray, “Trust in automation. Part II. Experimental studies of trust and human intervention in a process control simulation,” *Ergonomics*, vol. 39, no. 3, p. 429, 1996.

[29] D. D. Woods, E. M. Roth, and K. B. Bennett, “Cognition, Computing, and Cooperation,” S. P. Robertson, W. W. Zachary, and J. B. Black, Eds. Norwood, NJ, USA: Ablex Publishing Corp., 1990, pp. 123–158.

[30] P. Brézillon, “Context in Human-Machine Problem Solving: A Survey,” *A Survey, Knowledge Engineering Review*, vol. 14, pp. 1–37, 1996.

[31] P. Brézillon and J.-C. Pomerol, “CONTEXTUAL KNOWLEDGE SHARING AND COOPERATION IN INTELLIGENT ASSISTANT SYSTEMS,” *Le Travail Humain*, vol. 62, no. 3, pp. 223–246, 1999.

[32] A. Dey, D. Salber, and G. Abowd, “A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications.” 2001.

[33] P. Dourish, “What we talk about when we talk about context,” *Personal Ubiquitous Comput.*, vol. 8, no. 1, pp. 19–30, 2004.

[34] F. C. Keil, “Explanation and Understanding,” *Annu Rev Psychol*, vol. 57, pp. 227–254, 2006.

[35] F. Heider, *The Psychology of Interpersonal Relations*. Psychology Press, 1958.

[36] A. C. Graesser and B. A. Olde, “How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down,” *Journal of Educational Psychology*, vol. 95, no. 3, pp. 524–536, 2003.

[37] T. Lombrozo, “The structure and function of explanations,” *Trends in Cognitive Sciences*, vol. 10, no. 10, pp. 464–470, Oct. 2006.

[38] W. C. Salmon, *Four Decades of Scientific Explanation*, 1 edition. University of Pittsburgh Press, 1989.

[39] C. G. Hempel and P. Oppenheim, “Studies in the Logic of Explanation,” *Philosophy of Science*, vol. 15, no. 2, pp. 135–175, 1948.

[40] M. Scriven, “Explanations, Predictions, and Laws,” in *In Scientific Explanation, Space, and Time (Minnesota Studies i n the Philosophy of Science:*, vol. Vol-3, H. Feigl and G. Maxwell, Eds. Minneapolis: University of Minnesota Press, 1962, pp. 170–230.

[41] G. L. Murphy and D. L. Medin, “The role of theories in conceptual coherence,” *Psychological Review*, vol. 92, no. 3, pp. 289–316, 1985.

[42] D. C. Dennett, *The Intentional Stance*, Revised ed. edition. Cambridge, Mass.: A Bradford Book, 1989.

[43] W. Bechtel and A. Abrahamsen, “Explanation: a mechanist alternative,” *Stud Hist Philos Biol Biomed Sci*, vol. 36, no. 2, pp. 421–441, Jun. 2005.

[44] “Mechanistic Explanation: Ontic vs Epistemic.” [Online]. Available: https://www.academia.edu/1718331/Mechanistic_Explanation_Ontic_vs_Epistemic.

[45] “Help! My car is allergic to vanilla ice cream; a study in problem solving. The complete case - Kepner-Tregoe.” [Online]. Available: http://www.kepner-tregoe.com/blog/help-my-car-is-allergic-to-vanilla-ice-cream-a-study-in-problem-solving-the-complete-case/.

[46] S. Mazac, “Approche décentralisée de l’apprentissage constructiviste et modélisation multi-agent du problème d’amorçage de l’apprentissage sensorimoteur en environnement continu. Application à l’intelligence ambiante.,” PhD thesis, Université de Lyon, 2015.

[47] A. Najjar and P. Reignier, “Constructivist Ambient Intelligent Agent for Smart Environments,” in *PerCom - IEEE International Conference on Pervasive Computing and Communications*, San Diego, United States, 2013.

[48] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, “How to grow a mind: Statistics, structure, and abstraction,” *science*, vol. 331, no. 6022, pp. 1279–1285, 2011.

[49] A. Gopnik and L. Schulz, “Mechanisms of theory formation in young children,” *Trends in Cognitive Sciences*, vol. 8, no. 8, pp. 371–377, Aug. 2004.

[50] N. D. Goodman, T. D. Ullman, and J. B. Tenenbaum, “Learning a theory of causality,” *Psychological Review*, vol. 118, no. 1, pp. 110–119, 2011.

[51] C. Kemp, A. Perfors, and J. B. Tenenbaum, “Learning overhypotheses with hierarchical Bayesian models,” *Dev Sci*, vol. 10, no. 3, pp. 307–321, May 2007.

[52] L. Scanu, S. Ploix, P. Bernaud, and E. Wurtz, “Model Tuning Approach For Energy Management Of Office and Apartment Settings,” in *Building Simulation 2017*, San Francisco, United States, 2017.

[53] N. S. Kim and F. C. Keil, “From symptoms to causes: Diversity effects in diagnostic reasoning,” *Memory & Cognition*, vol. 31, no. 1, pp. 155–165, 2003.

[54] L. J. Rips, “Circular reasoning,” *Cognitive Science*, vol. 26, no. 6, pp. 767–795, Nov. 2002.

[55] R. S. Nickerson, “The Projective Way of Knowing: A Useful Heuristic That Sometimes Misleads,” *Current Directions in Psychological Science*, vol. 10, no. 5, pp. 168–172, Oct. 2001.

[56] W. Whewell, *The Philosophy of the Inductive Sciences, Founded upon their History*. London, J. W. Parker, 1847.

[57] L. Rozenblit and F. Keil, “The misunderstood limits of folk science: an illusion of explanatory depth,” *Cognitive Science*, vol. 26, no. 5, pp. 521–562, Sep. 2002.

[58] J. Pearl, “The Art and Science of Cause and Effect,” *Causality*, 1996. [Online]. Available: http://bayes.cs.ucla.edu/BOOK-2K/causality2-epilogue.pdf.

[59] M. PAL *et al.*, “Analysis of Optimizers to Regulate Occupant’s Actions for Building Energy Management,” in *ICAPR 2017 - Ninth International Conference on Advances in Pattern Recognition*, Bangalore, India, 2017.

[60] E. Reiter and R. Dale, “Building Applied Natural Language Generation Systems,” *Nat. Lang. Eng.*, vol. 3, no. 1, pp. 57–87, Mar. 1997.

[61] F. Portet *et al.*, “Automatic Generation of Textual Summaries from Neonatal Intensive Care Data,” *Artificial Intelligence*, vol. 173, no. 7–8, pp. 789–816, 2009.

[62] N. McIntyre and M. Lapata, “Learning to Tell Tales: A Data-driven Approach to Story Generation,” in *ACL 2009, Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the AFNLP, 2–7 August 2009, Singapore*, 2009, pp. 217–225.

[63] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, “Show and tell: A neural image caption generator,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[64] R. (1982) Boitet C., “‘DSE-1’— Le point sur ARIANE-78 début 1982. Contrat ADI/CAP-Sogeti/Champollion (3 vol.), GETA, Grenoble, janvier 1982, 616 p.”

[65] A. A. Alyafi, J. Guillbaud, P. Reignier, and S. Ploix, “Explanations engine for energy management systems in buildings,” in *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 2017, vol. 2, pp. 722–729.

[66] J. Rowan and E. D. Mynatt, “Digital Family Portrait Field Trial: Support for Aging in Place,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2005, pp. 521–530.

[67] K. Hornbaek, B. B. Bederson, and C. Plaisant, “Navigation Patterns and Usability of Zoomable User Interfaces with and Without an Overview,” *ACM Trans. Comput.-Hum. Interact.*, vol. 9, no. 4, pp. 362–389, Dec. 2002.

[68] Y. Laurillau *et al.*, “The TOP-Sliders for multi-criteria decision making by non-specialists,” presented at the NordiCHI, Oslo, 2018.

[69] A. F. Blackwell and T. R. Green, “Investment of attention as an analytic approach to cognitive dimensions,” in *Collected Papers of the 11th Annual Workshop of the Psychology of Programming Interest Group (PPIG-11)*, 1999, pp. 24–35.

