IASDR 2019
International Association of Societies of
Design Research Conference 2019
DESIGN REVOLUTIONS

Manchester School of Art
Manchester Metropolitan University
02-05 September 2019

Anticipatory design: A future-led mixed-methodology to mitigate unintended consequences

Galdon, Fernando [a]; Hall, Ashley [a]; Wang, Stephen J. [a]

[a] Royal College of Art, London, UK
* fernando.galdon@network.rca.ac.uk

In this paper, the authors proposes anticipatory design as a future-led mixed-method to address unintended consequences. It combines systems analysis with extrapolations and constructivist perspectives to reconcile confronted models of design future(s). In the results presented, the authors suggest a need to include ethical frameworks in design to involve students in ethical issues to address the main task of design in the digital and exponential technological age within which we are living including; preparedness, readiness, and appropriateness.

Keywords: design futures; preparedness ; readiness; appropriateness; anticipation

1. Introduction

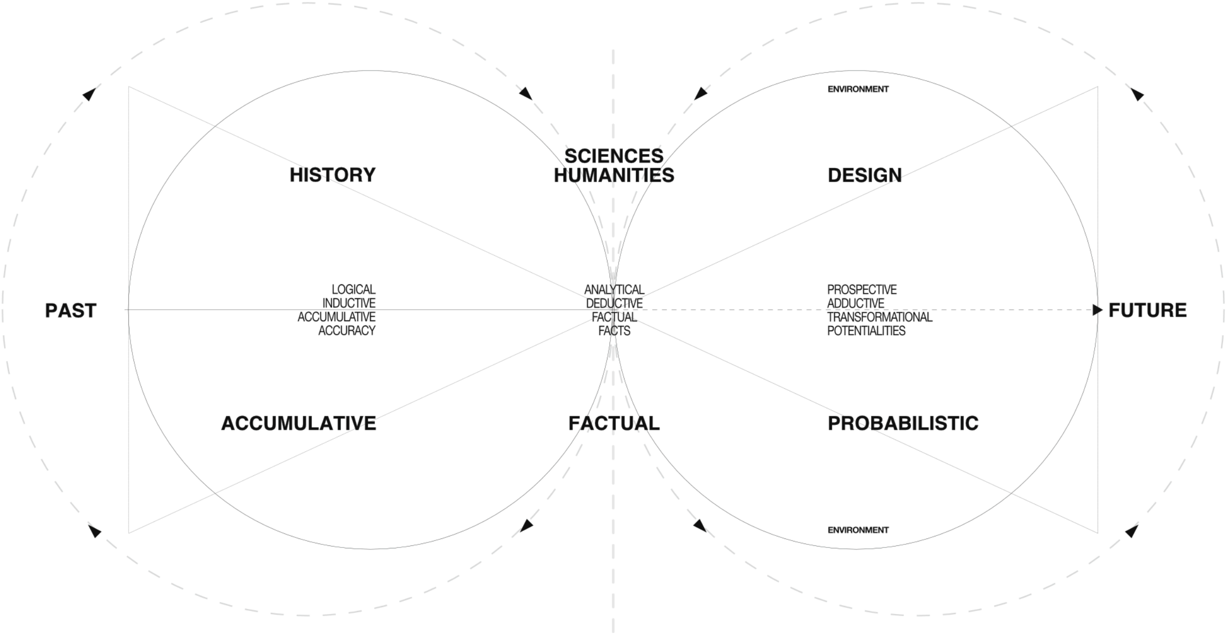
1.1. Anticipation and design

As are moving from the industrial to the digital age, the acceleration of innovation is transforming reality and affecting the development of society and the nature of design practice. In this context, recent strategies in the social sphere call for anticipatory strategies. For instance, Guston introduced the idea of anticipatory governance defining it as:

> "…a broad-based capacity extended throughout society that can act on a variety of inputs to manage emerging knowledge-based technologies while such management is still possible." (Guston, 2014, p.??).

In a report presented by the Institute for the Future on 'anticipatory governance' (Future, 2009), the authors aim for processes that involve the simulation of possible futures to address anticipation as a strategy for good government.

Historically anticipating and designing the future has always been a human characteristic. In antiquity (1000BC - 1400AC), prophecies and alternative presents were introduced by priests and Greek and Roman philosophers such as Plato (The Republic) or Cicero. In the Renaissance (1400 - 1800), planetary explorations via utopias of other places were structured around mathematical and philosophical endeavours by the likes of Da Vinci or Thomas More. With the scientific revolution (1600 - 1700), observations became the main method to lucubrate biological and scientific-based futures with the likes of Bacon or Newton. In the Enlightenment (1700 -1900), theories of progress via theoretical and metaphysical insights became the main approach to construct the future. Finally, with the theories of Einstein and the integration of time directionality a clear notion of the future became settled. It led the transformational industrial era (1900 - 2000) where Knowledge-based futures were built via scientific, social and critical approaches. in 1927 Richard



Buckminster Fuller called for an 'industrially realisable design science’ (Fuller, 1992) through his ‘Eight strategies for a comprehensive anticipatory design science’. However this failed to fully materialise as a new field. Now with the advent of the digital age, accelerating technology complexity, black box technologies and wicked problems new prospective approaches are required to deal with the exponential nature of our emerging new era.

1.2. Framing design

One of the first design science theorist, John Chris Jones, postulated in the 1970s in his seminal book, Design method, that design was different from the arts, sciences, and mathematics. In response to the question "Is designing an art, a science or a form of mathematics?" Jones responded:

> ‘The main point of difference is that of timing. Both artists and scientists operate on the physical world as it exists in the present (whether it is real or symbolic), while mathematicians operate on abstract relationships that are independent of historical time. Designers, on the other hand, are forever bound to treat as real that which exists only in an imagined future and have to specify ways in which the foreseen thing can be made to exist.’ (Jones, 1992. pp. 10)

From this perspective, we would position design as a prospective thinking activity in the context of abductive reasoning (making decisions without having all the information) (Douven, 2011). In this area, research by Dorst (Dorst, 2010) or more recently Cramer-Petersen et al. (Cramer-Petersen, 2018) have concluded that design combines deductive and abductive reasoning, however, in both cases, abductive reasoning plays a fundamental role as initiator of the design activity. Furthermore, as the digital paradigm, with its exponential development (Kurzweil, 2005) and network uncertainty becomes more prevalent in design, practice will need to focus more in the preventive/anticipatory aspects of design (preparedness, readiness and, appropriateness). In this context, the deductive becomes limited by access and the abductive reasoning aspects becomes more dominant, prevalent and necessary.

This intrinsic prospective approach of design, based on abductive reasoning, planning, solution-based problem solving, problem shaping, synthesis, preparedness, readiness and appropriateness in the built environment determines a different model of knowing. In this scenario, the designer is dealing with wicked problems by accessing areas yet-to-be or not-fully-formed (Rittel & Webber, 1973; Buchanan, 1992; Conklin, 2006) . Consequently, its output is based on potentialities, not certainties. As Glanville proposed, ’knowledge for’ future action and possibilities rather than ‘knowledge of’ past actions and events (Glanville, 2005). In this context, as the life of the intervention is extended into the future, time to assess the impact of the design is extended during its lifetime and forever bounded to its environment. By exchange Validation therefore is always a posteriori, and the proposed output becomes the main element to be assessed. This intrinsically means that knowledge in design is probabilistic in nature. Design implies a posteriori development based on exchange which demands to go beyond existing time with a very clear function in mind; to transform.

Figure 1. Timeframe model. Source: Fernando Galdon.

Within this scenario, an investigative overview of twentieth century approaches to future studies structured prospective design practises around two main approaches; the scientific-positivistic based on the method of extrapolation (1900-1950), and a sociological-pluralistic perspective based on constructivism (1950-2010).

1.3. Designing the future

1.3.1. Scientific Empirical
Methods based on Newtonian physics. This approach is based on the systematic practice of repeating laboratory experiments and controlling variables to establish proof of our hypothesis. Main methods: extrapolations of historical data, utilisation of analytical models and the systematic use of experts as forecasters of opinion. This approach uses techniques based on Mathematics, Modelling, Simulation and, Gaming.

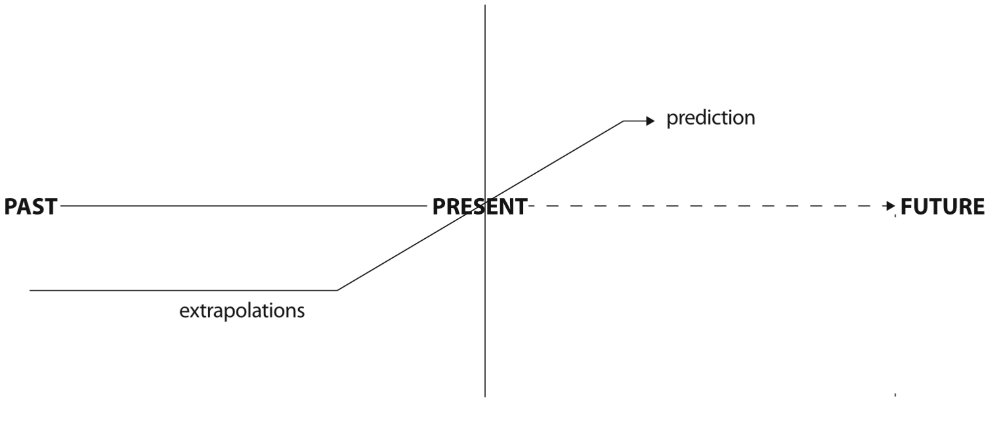


Figure 2. Positivistic model based on extrapolation. Source: Fernando Galdon.

1.3.2. Pluralistic Human-Centred

Methods based on sociology. This approach is based on the social and critical practice of mapping a wealth of possibles futures. Main methods: contextual data analysis, interpretative analytical methods and the systematic use of participatory methods. This approach uses techniques based on Cones, Mind maps, Future wheels and, Flow-scapes.

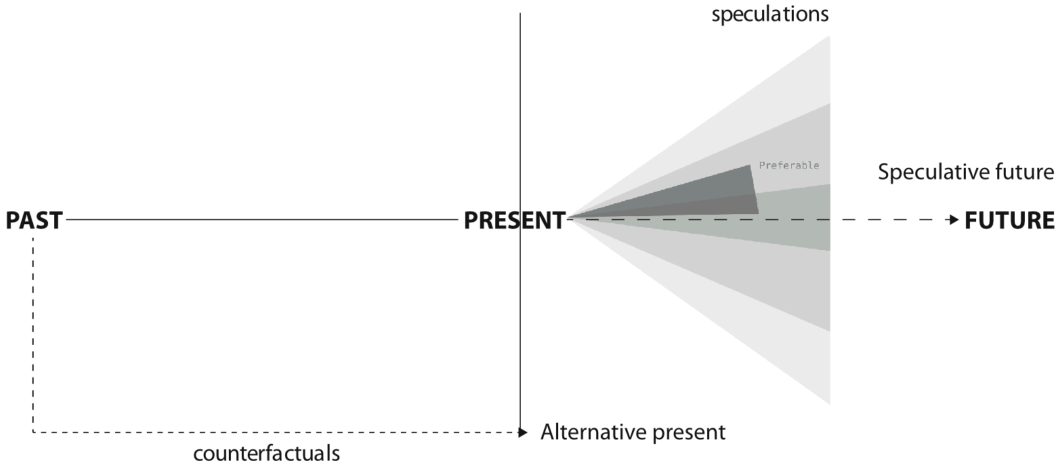


Figure 3. Pluralistic model based on constructivism. Source: Bezold, C. and Hancock, T. (1994), Voros, J. (2003), and Auger, J. (2012)

1.3.3. Critical analysis

Although these perspectives have been widely used, they present limitations. The scientific/ positivistic approach is perceived as objective and values-neutral, however, it is also perceived as presenting narrowness in focus (only one possible future) and lack of contextual awareness. On the other hand, the pluralistic approach is perceived as inclusive and impartial, however, it is also perceived as presenting a loose focus (too many possible futures) and is too dependent of contextual awareness (Gidley, 2017).

In terms of the widest used methodology of speculative design, one of the fundamental advantages is that it removes a range of constraints normally used in product design. It limits the validity of its outcome to plausibility and the uncanny (Auger, 2012). However, it creates a lateral problem; the difficulty of controlling the speculation. As a result, many of the proposed outputs end in what future studies expert Jennifer Gidley names ‘Pop futurism’ (superficial and media-friendly outputs) (Gidley, 2017).

In this paper, the authors consider both limitations and propose a mixed-methodology aimed at enhancing the positive side of each confronted approach and present an integrative model aimed to reconcile different perspective to improve the main task of design in the digital and exponential technological age we are living in via; preparedness, readiness and, appropriateness.

2. Method

The methodological approach we have used includes literature reviews and research through design to develop a proposed model. Academic conferences were used to validate the model. Finally, workshops and co-design activities were implemented to evaluate key elements of the proposed model.

Literature reviews focused on future studies, design futures and on models of design research. Research through design was implemented in the sense of using the design process as a critical and reflective tool to investigate limits and opportunities in the design discipline to develop potential methods and techniques. In the process, it uses system analysis to underpin a potential case study on virtual assistants to develop the intended framework. In this context, academic conferences were targeted to validate different aspects of the proposed case. Finally, as design is not a linear process and depends on emergent elements, iterative evaluations were conducted via two co-design workshops on the relationship between design and futures at the Royal College of Art to test the core aspects of the proposed framework.

3. Discussion

3.1 Anticipatory design model development

3.1.1 Trajectories

First, building from the literature review, the leading author used timelines as graphical projective tools to gain a contextual understanding of the technology at hand and project a possible trajectory based on relational patterns. The main author approached its design mainly by dividing the space into two equal parts by drawing the timeframe in the middle. This action immediately created two spaces which were used as comparative or relational spaces for prospective inquiry and analysis aiming to spatialise abductive thinking. In total, two timelines were implemented. First, the author implemented a contextual analysis of the system to generate a hypothesis. Then, a second timeline was implemented to underpin a case study to address the initial hypothesis.

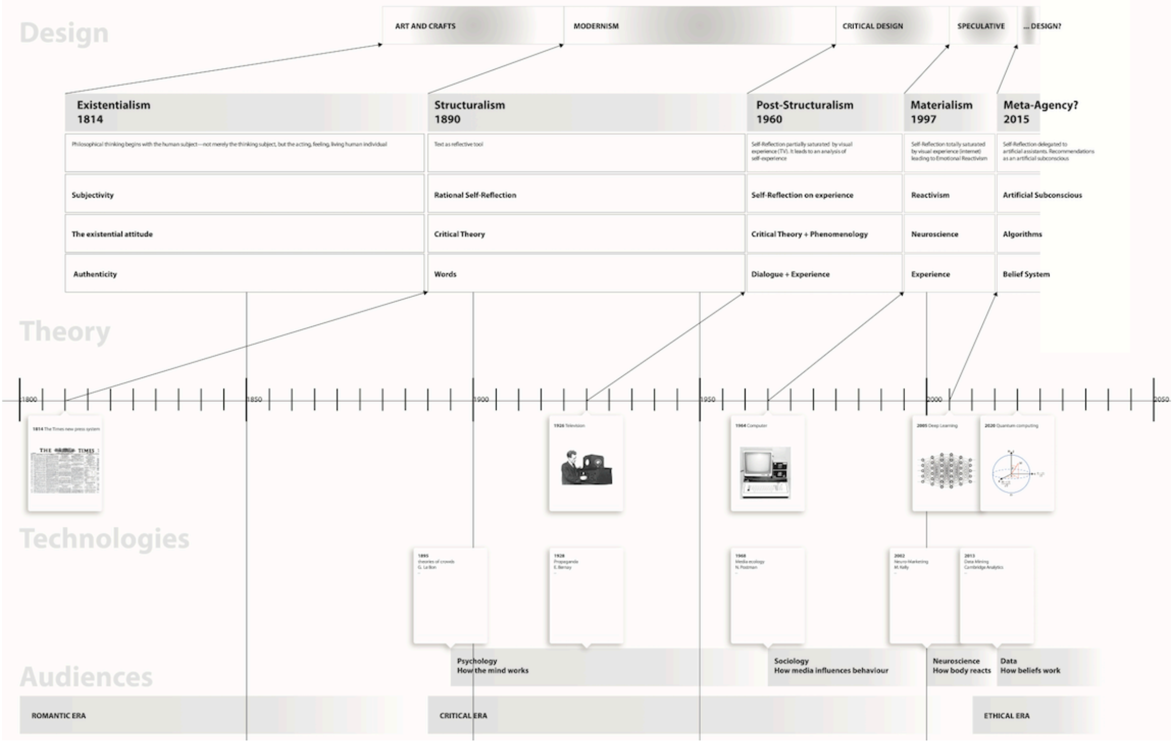


Figure 4. Understanding relational patterns among technology, theory and practise for prospective analysis. Source: Fernando Galdon.

The first timeline focused on the relationships between technology, theory and practice. It underlined a range of impactful elements based on the potential impact of AI; the emergence of meta-agency, the emergence of an artificial subconscious, the relevance of algorithms and the impact of belief systems. These elements led to building a hypothesis around Virtual Assistants, and the potential need for a new kind of design to address all these elements.

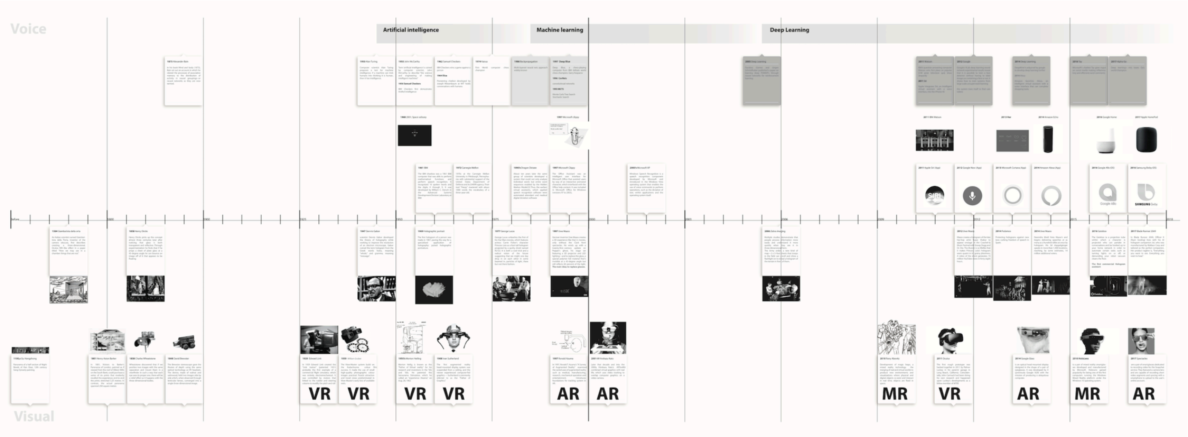


Figure 5. System-based relational analysis in virtual assistants. Source: Fernando Galdon.

Once a case study was underpinned, a second timeline was implemented to understand the context of Virtual Assistants. This systems-based relational analysis presented the key technology of Natural Language Processing (NLP) and its embodying potentialities (robots and holograms) as the main elements to address.

3.1.2. Probabilistic extrapolations

As we are projecting the interaction into the future, questions of evidence regarding the prospective development and impact of emerging technology raised. In this context, due to the limited access of emerging technologies by researchers, three elements were used to underpin probabilistic extrapolations;

- Demos: Demos are introduced by tech companies to illustrate the potentialities of new technologies. They can be used by researchers to understand the potential development of emerging technologies. In this case, the author selected a demo called Duplex introduced by Google. The extraordinary levels of fluidity, coherence, and autonomy presented a case to understand the evolutive nature of Virtual Assistants form queries to conversations and from reactive to proactive interactions.

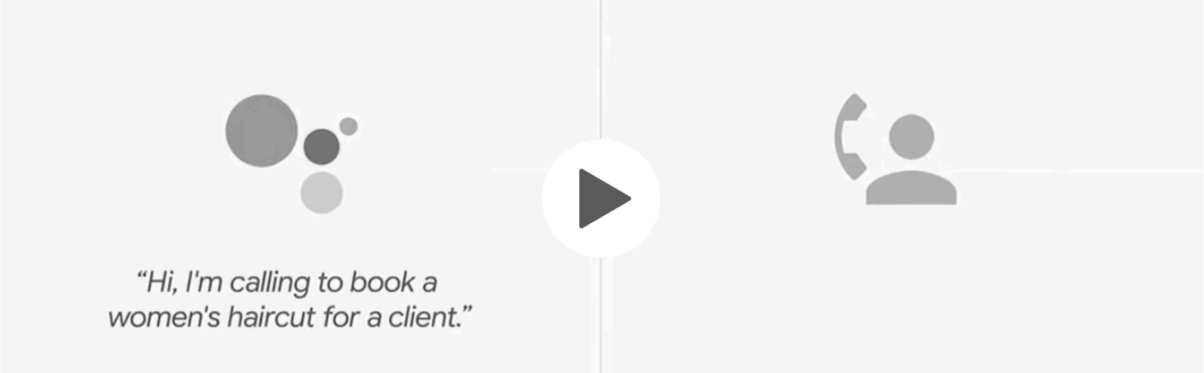


Figure 6. Duplex demo by Google. Source: Google.

- Prototypes; Prototypes also present a case on potential technological developments. As an example, the author conducted research into state of the art technology and underlined a prototype capable of predicting depression. This prototype raised ethical questions and illustrates how technology may impact our lives in a positive or negative manner. (Eichstaedt, *et al.*, 2018)

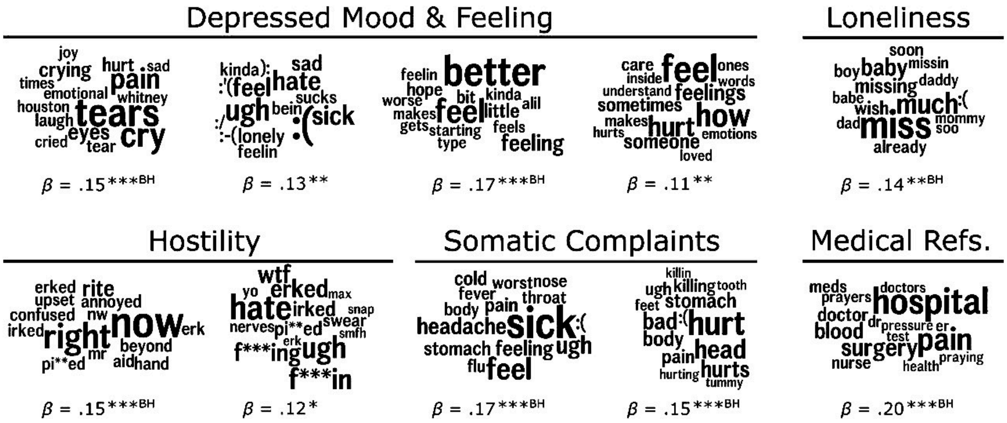


Figure 7. Depression prediction algorithm. Source: (Eichstaedt, et al. 2018)

- Patents; Patents also illustrate the potential development of a given technology. As an example, the author conducted research into patent applications to underpin potential developments in the context of Virtual Assistants. A clear case was a patent filed by Amazon capable of diagnosing a cough and providing treatment. This patent aims to transform Alexa into a doctor and raises many ethical questions regarding its implementation (Jin, 2018).

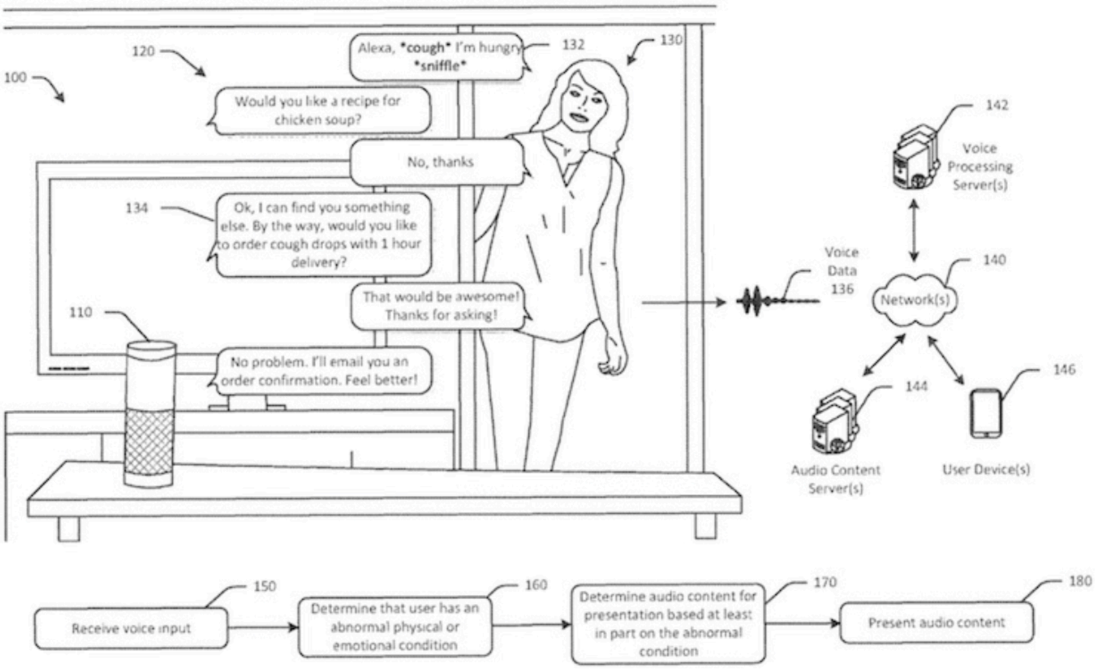

100
120
Alexa, *cough* I'm hungry *sniffle*
132
130
Would you like a recipe for chicken soup?
No, thanks
134
Ok, I can find you something else. By the way, would you like to order cough drops with 1 hour delivery?
That would be awesome! Thanks for asking!
110
No problem. I'll email you an order confirmation. Feel better!
142
Voice Processing Server(s)
Voice Data 136
140
Network(s)
144
146
Audio Content Server(s)
User Device(s)
150
Receive voice input
160
Determine that user has an abnormal physical or emotional condition
170
Determine audio content for presentation based at least in part on the abnormal condition
180
Present audio content

Figure 8. Cough prediction algorithm patent. Source: Amazon (Jin, 2018)

These examples illustrate how designers can use these elements - demos, prototypes and, patents - to anticipate potential positives and/or negative interactions.

3.1.3. Asymmetries

In order to understand the positive and negative potential dynamics of the system asymmetries needed to be understood and identified. They uncovered potential areas of conflict, exploitation and injustice which may have a tremendous impact on society and businesses. As an example, building from a case study on Facebook and Cambridge Analytica, data asymmetry became a major element to address. Therefore positioning this process as key for the successful development of the project.

3.1.4. Consequences

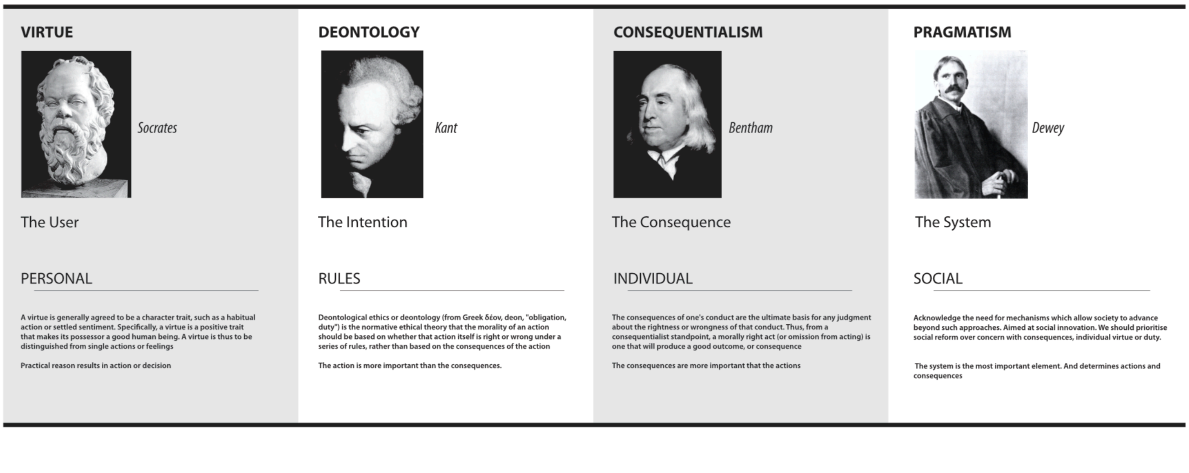
This area aims to integrate ethical analysis into the development of new products and services. Ethics focuses on how a person should behave. It is a philosophy applicable to daily life or existence. It integrates two areas in order to determine rules or codes of conduct; philosophy, the art as asking questions, and morality, what is good or bad. Its main objective is to determine the right thing to do. Its ontology is based on creating social constructs for the adequate functioning of society. It's epistemology to decode these constructs while its output aims to set standards of behaviour for daily life. Once the area was defined, a literature review on normative ethical frameworks was conducted. From this process, a debate emerged on which framework to use; Socrates's virtue, Jeremy Bentham's Consequentialism, Emmanuel Kant's Deontology or John Dewey's Pragmatism.

Virtue refers to being. In this paradigm, morality emerges from the identity of the individual rather than their actions or consequences. Socrates approach refers to an end to be sought. It asserts that the right action will be that chosen by a suitably 'virtuous' agent. Practical reason results in action or decision.

Consequentialism states that the consequences of somebody actions are the ultimate basis for any kind of judgment regarding that action. This perspective is non-descriptive, in the sense that the value of the action is determined by its consequences rather than its intentionality. It focuses on the outcome of conduct.

In deontology, the rightness or wrongness of actions does not depend on their consequences but on whether they fulfil our duty or not. These actions are conditioned by a set of rules, may they be natural, religious or social.

Pragmatism aims for social reform as a strategy to address morality. Actions and consequences are possible because the context or system allows for them. Aimed at social innovation, in this perspective we should prioritise social reform over concerns with consequences, individual virtue or duty.



A virtue is generally agreed to be a character trait, such as a habitual action or settled sentiment. Specifically, a virtue is a positive trait that makes its possessor a good human being. A virtue is thus to be distinguished from single actions or feelings

Practical reason results in action or decision

Deontological ethics or deontology (from Greek δέον, deon, "obligation, duty") is the normative ethical theory that the morality of an action should be based on whether that action itself is right or wrong under a series of rules, rather than based on the consequences of the action

The action is more important than the consequences.

The consequences of one's conduct are the ultimate basis for any judgment about the rightness or wrongness of that conduct. Thus, from a consequentialist standpoint, a morally right act (or omission from acting) is one that will produce a good outcome, or consequence

The consequences are more important that the actions

Acknowledge the need for mechanisms which allow society to advance beyond such approaches. Aimed at social innovation. We should prioritise social reform over concern with consequences, individual virtue or duty.

The system is the most important element. And determines actions and consequences

Figure 9. Normative ethics main frameworks. Source: Fernando Galdon

The fundamental problem with Dewey's perspective is that in order to change the system, we need an alternative or global consensus. As illustrated by Professor Harari, AI is a global problem such as climate change or nuclear war which entails global consensus (Harari, 2019). Insofar as we have not reached this consensus it is not an adequate framework to address the design of a system.

In Socrates virtue, the fundamental problem is the limited capability of humans to assess what is happening. The acceleration and volume of information delivered by social interactions and algorithmic updates is fragmenting reflection and cognition by disconnecting the pre-frontal cortex by saturation; our attention span has been reduced from 12" to 8" in four years by multitasking (National Center for Biotechnology Information, 2016) (Kahneman, 2011) and after 21 minutes comparing information our pre-frontal cortex shuts down (Mullins, 2013) and only information with a big emotional impact is retained (Buchanan, 2007). These processes are transforming society from reflective to reactive. The digital era is bringing Emotional Reactivism as its main paradigm. It is questioning the idea of truth and reality and repositioning the decision centre from reason to emotional experience. Thus invalidating the model proposed by Socrates based on reason.

In this scenario, two main candidates remain. On one side, Jeremy Bethan's consequentialism. On the other, Emmanuel Kant's deontology. The former situates the ethical intervention on the consequence, whereas, the latest, on the intentionality. In terms of deontology, fundamental problems are interpretability and interruptibility. The system does not know what is doing, therefore, it cannot stop. According to researchers from the most advanced AI company in the world DeepMind, this is currently impossible (Ortega, 2018). Insofar as we are not capable of designing them, it is not a suitable strategy. Consequently, the only paradigm remaining is Consequentialism. In this framework the fundamental elements are the consequences of an action, therefore, the system will be judged by the consequences of its actions.

In this scenario, a design framework-toolkit presented by Mark Michael to address unintended consequences was integrated into the design process (Michael, 2019). However, it proved limited as contexts and actions emerged from the literature as fundamental variables to address (Bradshaw, 2013). These elements became integrated via the design of a multi-focus system analysis process capable of integrating different perspectives.

3.1.5. Counter-fictions

Counter-fiction is an experimental emerging area in design practice. So far, only two publications were found during this research that explore its possibilities; A monographic journal issue (Multitudes, 2012), and a book (Belliot, 2018). This approach aims to address the relations of domination. Its main approach, rather than being imposed or forced, is based on the co-production of control systems aimed to decrease repression and enhance individual freedom and responsibility. In this paradigm:

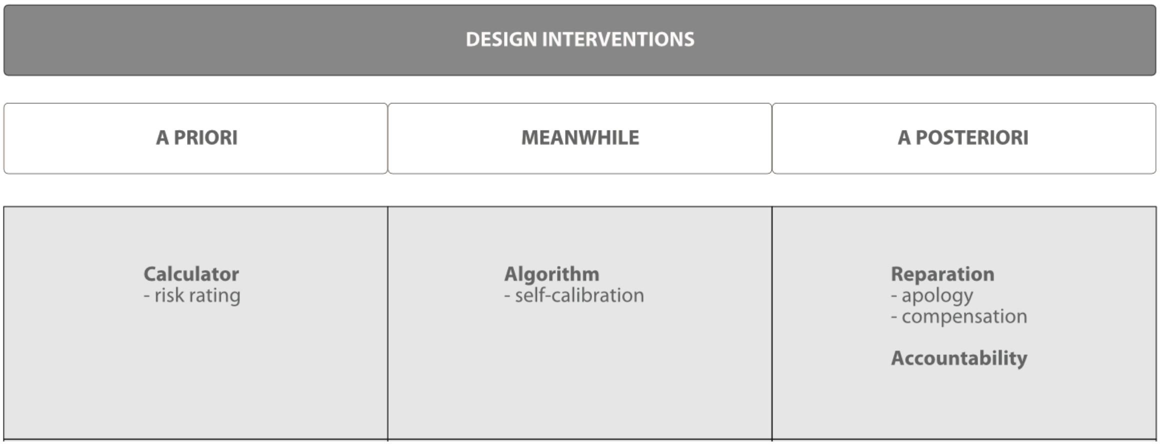
> 'Freedom is nothing other than the correlative of the implementation of security devices. A form of power announced as "near future" or immediate present, which makes obsolete old forms of resistance still indexed on disciplines and forces us to invent "new weapons'' (Foucault on Claisse, 2012, pp.108)

Control is the main element to account for. It is understood as a mode of relationships between individuals. In this relational perspective, power is a dynamic and reciprocal force addressed through asymmetric relations in which the controlled one sees his actions, cognitions and possible effects reduced, although not totally determined by the controller. Power can be seen as a relation or as an influence, and differs from the point of view of the spectrum of possibilities actually controlled by individuals. This approach places trust as a fundamental variable to build and maintain the relationship.

In this context, the use of counter-fictional strategies emerged for the author as a strategy to address the dynamics of the system, but also as an experimental method to ground speculations. Its intervention can be placed a priori, meanwhile or a posteriori.

Building from a literature review, the author underpinned levels of automation (LoA) as a tool to address trust in automated systems. Gradient-based models of approximation have been used extensively in the field to address trust in automated systems. This approach has been consistent in the automation literature since its introduction by Sheridan and Verplanck (1978). Levels of automation (LoA) is acknowledged by Kaber (2018) as a fundamental design characteristic that determines the ability of operators to provide effective oversight and interaction with systems autonomy. In this context, a preliminary level of automation was built. However, contexts and actions emerged as capital variables to address two fundamental questions; if something goes wrong, how can we repair trust in the system? and, Who should be accountable for the reparation?

In this scenario reparation raised as an element to address. This acknowledgment led to the articulation of two complementary scales; levels of reparation and levels of accountability. These two scales became a posteriori design intervention. At the same time, by unifying these scales with the automation scale and the variables of contexts and actions, and the integration of the variables around access; a calculator was build to generate a trust rating by which to understand the risk of a particular action. This design became an a priori design intervention. Finally, by combining a priori and a posteriori interventions an algorithm could be designed to allow the system to self-calibrate. This intervention becomes a meanwhile design intervention.



DESIGN INTERVENTIONS		
A PRIORI	MEANWHILE	A POSTERIORI
Calculator - risk rating	**Algorithm** - self-calibration	**Reparation** - apology - compensation **Accountability**

Figure 10. Design interventions. Source: Fernando Galdon.

3.1.6. Final model - Behavioural

Relational methods based on ethics. This approach is based on the systematic practice of relational system analysis to predict and model behaviour. Main methods: historical data analysis, relational frameworks and the systematic use of ethical methods. This approach uses techniques based on Trajectories , Probabilistic extrapolations, Asymmetries, Consequences, and Counter-fictions

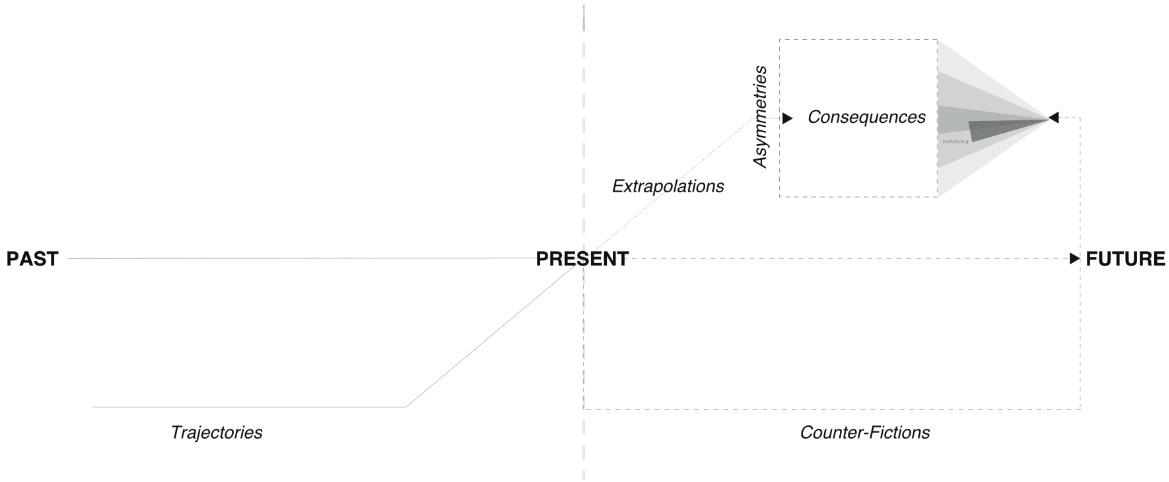
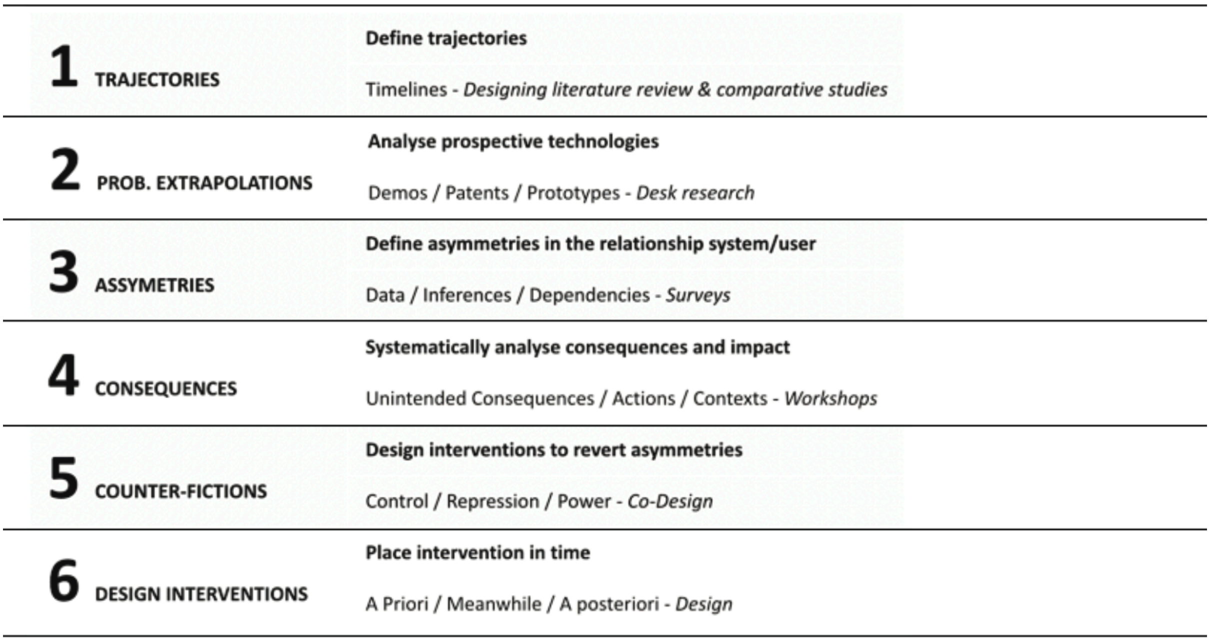


Figure 11. Anticipatory model based on systems analysis. Source: Fernando Galdon.



1 TRAJECTORIES	**Define trajectories** Timelines - *Designing literature review & comparative studies*
2 PROB. EXTRAPOLATIONS	**Analyse prospective technologies** Demos / Patents / Prototypes - *Desk research*
3 ASSYMETRIES	**Define asymmetries in the relationship system/user** Data / Inferences / Dependencies - *Surveys*
4 CONSEQUENCES	**Systematically analyse consequences and impact** Unintended Consequences / Actions / Contexts - *Workshops*
5 COUNTER-FICTIONS	**Design interventions to revert asymmetries** Control / Repression / Power - *Co-Design*
6 DESIGN INTERVENTIONS	**Place intervention in time** A Priori / Meanwhile / A posteriori - *Design*

Figure 12. Anticipatory design methods description and interventions. Source: Fernando Galdon.

3.2 Model test

3.2.1 Academia

The levels of automation, reparation, and accountability were tested via a survey and five papers were produced. The foundational paper of levels of automation will be presented and published in the proceedings of INAIT'19 at the University of Cambridge (Galdon, 2019a). The levels of reparation and accountability will be presented and published in the proceedings of IHIET'19 at University of Côte d'Azhur (Galdon, 2019b), (Galdon, 2019c). The calculator has been presented and published in the proceedings of MIT A+B Applied Engineering conference at the Massachusetts Institute of Technology in May 2019 via an applied case on user engagement optimisation to enhance energy consumption and management (Galdon, 2019d). In this conference, the author was also invited to present an additional poster illustrating the research through design process (Galdon, 2019e). Finally, a collaborative project is being discussed to develop a proof-of-concept for the self-calibrating algorithm.

3.2.2 Co-design workshops

The first Workshop invited 20 participants from the School of Design at the RCA to test on the first hand, differences amongst group and individual work, and on the other hand, the simplified systematic analysis of unintended consequences presented by Mark Michael (Michael, 2019). The participants were distributed in four groups of five members.

Giving a potential technological development, the framework presented by Michael demanded participants to analyse four elements; anticipated desired, anticipated undesired, unanticipated desired and unanticipated undesired potential outputs. As a result, the anticipated quadrants were better developed with 61 proposals, whereas the unanticipated aspects of product development presented 54 proposals in overall from the participants. Unanticipated undesired outcomes presented a very clear challenge for participants. They were referential to known issues. Answers were logical, rational and expected. There was a lack of originality and incapability to go ‘beyond’. The main author had to instigate debate by introducing some examples. However, instead of opening the scope of outputs, these examples become replicated by variation or integration. Occasionally, some participants proposed interesting ideas, but the group dynamics demanded consensus and prevented them going ‘beyond’ what they already knew, thus limiting abductive thinking and jeopardising anticipatory strategies. In anticipatory contexts is fundamental to go ‘beyond’. Only if you can imagine contentious developments, you can develop strategies to mitigate prospective consequences.

The second hour in the first workshop aimed to redo the same task from an individual perspective. A booklet for individual development was distributed among participants. The engagement was articulated around the idea that they could re-appropriate the method by integrating their own individual research into the process. Half an hour into the task and half of the participants left the workshop. It seems that they need constant engagement, and when requested to conduct individual work and reflect within themselves, they tended to disengage and abandon the task. The other half engaged as expected, with 20% of participants engaging vigorously, to the extent of asking whether they could carry the task at their homes after the workshop. Yet, outcomes were building from the previous task. Again, a lack of ‘going beyond’ was present. The analytical model used presented clear limitations.

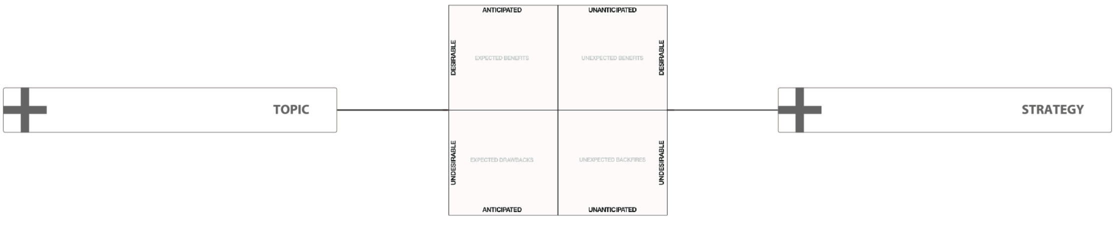


Figure 13. Consequential analysis. Source: Fernando Galdon from Mark Michael.

The second workshop invited 10 participants from the School of Design and Architecture at the RCA to test and improve a multi-layered approach to systematically analyse consequences by addressing contexts and actions to propose mitigating strategies. Participants were distributed in two groups. The workshop was structured completely to operate as a group task to maintain engagement. All the participants completed the 2 hours workshop and they engaged consistently through all the stages.

The second workshop aimed to further investigate anticipatory analytical skills. As a result, the author introduced a range of variations. First, students mapped the current state of the art. (what a virtual assistant can do today). Then, in order to address originality and lack of ‘going beyond’, it introduced a What if …? approach to allow participants to break the logical and rational thinking and project possible or potential developments of the technology. This task was successful and unexpected outcomes emerged, allowing participants to go ‘beyond’ what already exists. This approach included positive and negative outcomes.

In terms of outputs, the workshop aimed to understand if speculative insight could be grounded by applying a systematic analysis between the insight and the design activity. The system analysis consisted of a three-level analytical process of the system at hand. First, they were requested to conduct the consequences quadrant used in Workshop 1, however differently, each group mapped the anticipated desired and undesired, and by confronting both groups the unanticipated emerged for each group. This element presented participants with their own limitations and enhanced self-criticality. Then, they mapped the prospective outcomes in terms of impact in contexts and impact of actions. This analytical step allowed them to understand contexts and actions impact on users. Finally, participants were requested to complete a design activity consisting of developing preventive strategies to the potentially negative interactions they had mapped. They were requested to use counter-fictional principles to transform the dystopic into real-world strategies that could be applied. The results were successful and presented strategies aiming to ground speculation into potential real-case interventions.

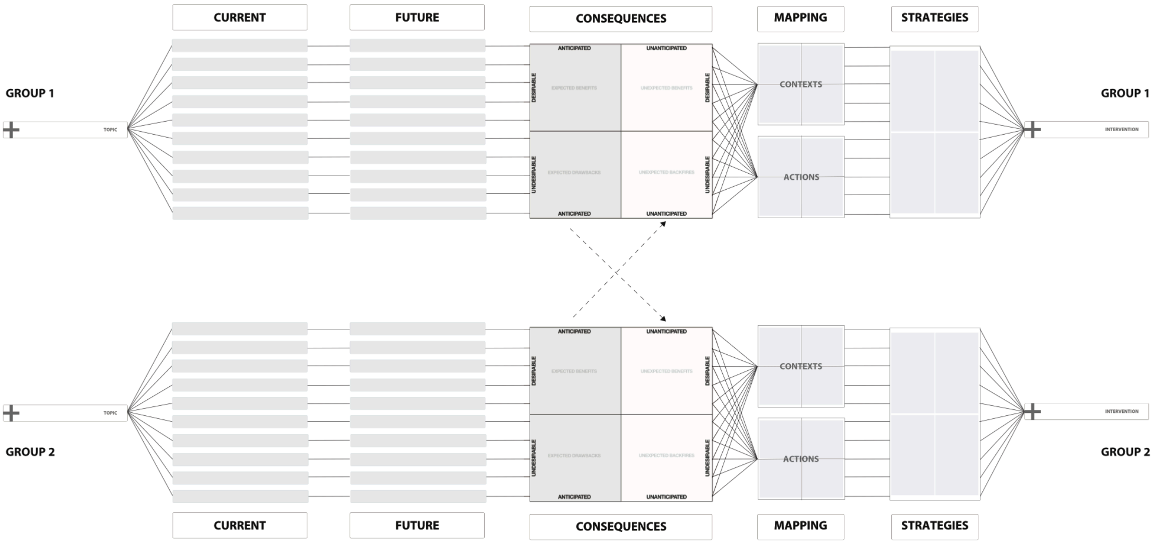


Figure 14. Multi-focus Consequential analysis for Anticipatory Design. Source: Fernando Galdon.

4. Conclusion

In this paper, the authors propose anticipatory design as a method to address unintended consequences. It combines systems analysis with extrapolations and constructivist perspectives to reconcile confronted models of design future(s).

In the results presented, the authors suggest a need to include ethical frameworks in design to involve students in ethical issues. To go beyond the positive impact of technology and design strategies to address and/or mitigate unintended consequences, as they are fundamental for the adequate development of society.

In developmental terms, results suggest that working in groups generates engagement, however, one of the fundamental problem of group tasks was that decisions were based on consensus when approaching the task from a rational and logical perspective, and some interesting ideas to address the potential impact of technological systems became superseded by the dynamics of the group. It recommends the integration of What if …? metaphysical affordances to break logical and rational analysis and enhance more distributed results. Furthermore, it is suggested the integration of a three-level consequential analysis including consequences, contexts, and actions to ground and focus the analysis. Finally, by implementing counter-fictional principles, results become real-world interventions aimed to address the main task of design in the digital and exponential technological age we are living; preparedness, readiness, and appropriateness to the build environment.

In the process, it challenges and evolves current notions in design research based on technological progress revolving around product development to a model based on ethical responsibility which places equal value on the process of design and the impact of the system in society. In this context, abductive thinking becomes the main design mindset in driving the transition from current to potential states leading to the mediation of anticipated and non-anticipated consequences. The anticipatory design framework introduces a process to deal with the increasing complexity of wicked problems, black box technologies and AI/ML technology acceleration, enhancing social values and ethical principles in the process.

This paper presents preliminary insights. Academic conferences and publications have been used to test the design outputs emerging from the process proposed and tailored workshops have tested key specific aspects of the methodology. Further research is being planned to test the full extension of the methodology proposed in educational and professional settings.

5. References

Auger, J. (2012). Why Robot? Speculative design, the domestication of technology and the considered future. PhD thesis. Royal College of Art.

Belliot, E. (2018). Counter-Fictional Design. Critique d’art[En ligne], Toutes les notes de lecture en ligne, mis en ligne le 04 novembre 2016, consulté le 01 novembre 2018. URL : http://journals.openedition.org/critiquedart/19220

Bezold, C. and Hancock, T. (1994). An Overview of the Health Futures Field. WHO Consultation, July 19-23

Bradshaw, J. M., Hoffman, R. R., Woods, D. D., & Johnson, M. (2013). The seven deadly myths of autonomous systems. IEEE Intelligent Systems, 28(3), 54–61.

Buchanan, R., (1992) Wicked Problems in Design Thinking, Design Issues, Vol. 8, No. 2, (Spring), pp. 5-21.

Buchanan, T. W. (2007). Retrieval of emotional memories. Psychological Bulletin, Vol 133(5), 761-779.

Claisse, F. (2012). Contr(ôl)e-fiction : de l'Empire à l'Interzone. Multitudes, 48(1), 106-117. doi: 10.3917/mult.048.0106. https://doi.org/10.3917/mult.048.0106.

Conklin, J., (2006) "Dialogue mapping." Building Shared Understanding of Wicked Problems. West Sussex, England: John Wiley & Sons.

Cramer-Petersen, C. L., Christensen, B. T., & Ahmed-Kristensen, S. (2019). Empirically Analysing Design Reasoning Patterns: Abductive-deductive Reasoning Patterns Dominate Design Idea Generation. Design Studies, 60, 39-70. DOI: 10.1016/j.destud.2018.10.001

Dorst, K. (2010). The nature of design thinking. DTRS8 Interpreting Design Thinking: Design Thinking Research Symposium Proceedings, 2010, pp. 131 - 139

Douven, I. (2011). ”Abduction", The Stanford Encyclopedia of Philosophy (Spring 2011 Edition), Edward N. Zalta ed., plato.stanford.edu/archives/spr2011/entries/abduction

Eichstaedt, J. C., Smith, R. J., Merchant, R. M., Ungar, L. H., Crutchley, P., Preoţiuc-Pietro, D., Asch, D. A., Schwartz, H. D. (2018). Facebook language predicts depression in medical records. Proceedings of the National Academy of Sciences Oct 2018, 115 (44) 11203-11208; DOI: 10.1073/pnas.1802331115

Foucault, M. (2004). Sécurité, Territoire, Population. Cours au Collège de France (1977-1978), Paris, Seuil, coll. « Hautes Études », p. 50. On Claisse, F. (2012). Contr(ôl)e-fiction : de l'Empire à l'Interzone. Multitudes, 48(1), 106-117. doi:10.3917/mult.048.0106. https://doi.org/10.3917/mult.048.0106.

Future, I. (2009). Anticipatory governance. Retrieved: 25 March 2019. Available from: http://www.iftf.org/ uploads/media/SR-1272_anticip_govern-1.pdf

Galdon, F., & Wang, S. J. (2019a). Designing trust in highly automated virtual assistants: A taxonomy of levels of autonomy. International Conference on Industry 4.0 and Artificial Intelligence Technologies. Cambridge, UK. ISBN: 978-1-912532-07-0

Galdon, F., & Wang, S. J. (2019b). From apology to compensation; A multi-level taxonomy of trust reparation for highly automated virtual assistants. Proceedings of the 1st International Conference on Human Interaction and Emerging Technologies (IHIET 2019) conference August 22-24, 2019, Nice, France.

Galdon, F., & Wang, S. J. (2019c). Addressing accountability in highly autonomous virtual assistants. Proceedings of the 1st International Conference on Human Interaction and Emerging Technologies (IHIET 2019) conference August 22-24, 2019, Nice, France.

Galdon, F., & Wang, S. J. (2019d). Optimising user engagement in highly automated virtual assistants to improve energy management and consumption. Proceedings of the 2019 Applied Energy Symposium AEAB Conference Proceedings, MIT. 22-24 May 2019.

Galdon, F., & Wang, S. J. (2019e). Future development of AI Virtual Assistants (VAs) in Energy management and consumption. Proceedings of the 2019 Applied Energy Symposium AEAB Conference Proceedings, MIT. 22-24 May 2019

Gidley, J. M. (2017). The future; A very short introduction. Oxford University Press. DOI: 10.1093/actrade/9780198735281.001.0001

Glanville, R. (2005). The Unthinkable Doctorate: Brussels, Design Prepositions. Cybernetics Research. American Society of Cybernetics, UK and Australia

Guston, D. H. (2014). Understanding ‘anticipatory governance.’ Social Studies of Science, 44(2), 218–242. https://doi.org/10.1177/0306312713508669

Harari, Y. N. (2019). Professor Yuval Noah Harari In conversation with Lord Hague of Richmond. RUSI, 13th November 2018. Available from: https://www.ynharari.com/wp-content/uploads/2018/12/20181113-RUSI-Harari_Discussion_TRANSCRIPT.pdf

Jin, H., Wang, S. (2018). Voice-based determination of physical and emotional characteristics of users. http://patft.uspto.gov/netacgi/nph-Parser?Sect1=PTO2&Sect2=HITOFF&u=%2Fnetahtml%2FPTO%2Fsearch-adv.htm&r=1&p=1&f=G&l=50&d=PTXT&S1=10,096,319&OS=10,096,319&RS=10,096,319

Jones, J. C. (1992). Design methods. New York: Van Nostrand Reinhold.

Kaber, D. B. (2018). Issues in Human–Automation Interaction Modeling: Presumptive Aspects of Frameworks of Types and Levels of Automation. Journal of Cognitive Engineering and Decision Making, 12(1), 7–24. doi:10.1177/1555343417737203

Kahneman, D. (2011). Thinking, fast and slow. New York: Farrar, Straus and Giroux.

Kurzweil, R. (2005). The Singularity is Near. New York: Viking Books. ISBN 978-0-670-03384-3.

Michael, M. (2019). An introduction to unintended consequences. Accessed; 16 March 2019. https://www.markmichael.io/insights/mapping-mitigating-unintended-consequences/

Mullins, P. a. (2013, November 22). Ground-Breaking project to brain-scan shoppers. Retrieved from Bangor University: https://www.bangor.ac.uk/news/university/ground-breaking-project-to- brain-scan-shoppers-16874

Multitudes (2012). Political Counter-Fictions – Fukushima: Voices of Rebels. No 48, 2012/1. Publisher : Assoc. Multitudes. ISBN : 9782916940779.

National Center for Biotechnology Information, U. N. (2016, July 2). attention span statistics. Retrieved from Statisticbrain: http://www.statisticbrain.com/attention-span-statistics/

Ortega, B. P. A. (2018). Building safe artificial intelligence : specification , robustness , and assurance Specification : design the purpose of the system. Medium. Retrieved from https://medium.com/@deepmindsafetyresearch/building-safe-artificial-intelligence-52f5f75058f

Rittel, H.W.J. & Webber, 1973. Dilemmas in a General Theory of Planning, M.M. Policy Sci 4: 155.

Sheridan, T. B., & Verplank, W. L. (1978). Human and Computer Control of Undersea Teleoperators: Fort Belvoir, VA: Defense Technical Information Center. https://doi.org/10.21236/ADA057655

Voros, J. (2003). A Generic Foresight Process Framework. Foresight, 5(3), pp.10-21

About the Authors:

Fernando Galdon: A Ph.D. candidate, Fernando is pursuing a doctoral programme in Global Innovation Design at the Royal college of Art, where He is investigating trust design at the intersection of Artificial Intelligence and society.

Ashley Hall: Ashley is Professor of Design Innovation at the Royal College of Art where he leads postgraduate research for the design school and the MRes in Healthcare Design. Ashley researches innovation methods, experimental design, design for safety, design pedagogy, globalisation design and cultural transfer.

Stephen Jia Wang: Head of Programme for the Global Innovation Design and Innovation Design Engineering programmes. Stephen explores challenges through international and/or interdisciplinary collaborative experimental design projects from the development of biomedical systems and devices, urban energy conservation systems, intelligent navigation systems, to the interactive exhibition media solutions.