Defence, Safety and Security
Kampweg 55
3769 DE Soesterberg
P.O. Box 23
3769 ZG Soesterberg
The Netherlands

www.tno.nl

T +31 88 866 15 00
F +31 34 635 39 77

TNO report
TNO 2022 R12321

Bias prevention in collaborative decision making

Date	January 2022
Author(s)	Dr J.E. Korteling, Dr ir M. Steen, Dr H.M.E Miedema
Number of pages	34
Project name	ERP Wise Policy Making
Project number	060.38077

Summary

Biases are systematic inclinations in human thinking and decision making occurring rather consistently in people and which may be counter-productive in several contexts. Mitigating these inherent thinking tendencies by training interventions or by building of expertise is difficult and will typically require lots of effort and time. Therefore, we argue that the most promising way of dealing with biases is to improve the *environment or context* in which people make decisions, instead of trying to directly improve their inherent thinking skills. One could stimulate or impose the use of certain working methods or aids with which the ingrained tendency towards biased thinking can be *prevented or circumvented*. This report explores the fundamentals of two possible behavioural methods of bias prevention, one relatively expensive and complete, one relatively cheap and pragmatic. Both methods focus on the circumvention of cognitive biases by improving the process of collaborative decision making in group dialogues, particularly in the context of policy making. The approaches are not mutually exclusive; they can be used simultaneously and complement each other.

The literature shows a multitude of, often similar and overlapping, biases, which makes it difficult to be aware and to counter all of them. The first approach is based on the idea that similar or overlapping biases may be addressed or prevented *as a group* by certain interventions in the decision making process or environment. Therefore, this approach starts with boiling down the multitude of biases into a limited number of ‘generic bias categories’. These bias categories are: ‘Status-quo and confirmational thinking’, ‘Retainment of information’, ‘Focusing and simplism’, ‘Short-term self-interest and in-group conformity’, ’Statistical blindness’. The identification of these five categories of cognitive biases makes it more feasible to develop a limited and well-arranged set of ‘debiasing working forms’. We expect that entire categories or groups of similar biases can be efficiently counteracted by using a these specific working methods. This report provides, for each category of biases, suggestions for feasible and adequate working forms and interventions.

A drawback of debiasing interventions may be that they may require quite some resources in terms of effort and time. In addition they require an experienced process facilitator, who must be able to adequately intervene and execute the working methods in interaction with the participants of the session. Therefore, we also propose a bias prevention approach that may be more pragmatic, intuitive, and simple. This method, which concerns a very first step towards an approach yet to be developed, builds on the use of ‘wise heuristics’. These are robust or generic principles that seem to have proven their value through time. An example of a wise heuristic may be: ‘Always start with a thorough inquiry of the question or problem to be solved’. This complementary approach aims at the use of these practical (and rational) principles. These wise heuristics approach, must be conceived as a first creative attempt to make an innovative step towards wise collaborative decision making. Therefore, in the near future we want to further develop, evaluate and improve this approach. The use of wise heuristics is a pragmatic and flexible technique for relatively quick collaborative decision making, e.g., with little resources and under time pressure. Based on both approaches, adaptive working methods can be introduced that may help to prevent biased thinking in collaborative decision making

Contents

1 Introduction: biases affect policy making

The present study is part of Early Research Project (ERP) Wise Policy Making of TNO. This ERP addresses the issue how to develop knowledge, insights, methods, and tools that harness wisdom[1] in policymaking. The ultimate goal of policymaking is to guide society towards a future of common human wellbeing. We consider this pursuit of wellbeing the ultimate goal of humanity (e.g., CPB-PBL-SCP, 2021; Harris, 2010). The issues we are confronted with nowadays are becoming increasingly complex and wicked, and concurrently the goals that we may choose to pursue are becoming more diverse and less clear. Although policymakers may be fully aware of this, they seem to lack the right tools and methods to effectively aim at obtaining long-term human wellbeing. In addition, many residents want more direct influence on important government decisions. So there is a growing need from society to be more involved in important policy choices, for example, about sustainability and climate. This involvement is made concrete by means of various forms of intensive citizen participation, such as citizens' forums. In a citizens' forum, a representative group of invited citizens develops policy proposals that can then be used at the administrative level (e.g., Advisory Committee on Citizen Involvement in Climate, 2021). In this way, the involvement of citizens can be increased, the support base for measures can be broadened and it is also possible to benefit from the knowledge in society. The main reason to involve citizens;' forums, however, is to make better decisions and better policies by bringing citizens' know-how and ideas together (Advisory Committee on Citizen Involvement in Climate, 2021). So, due to the growing intrinsic complexity of the challenges that our leaders and policy makers are confronted with, methods and tools will be increasingly indispensable to support them in making better collaborate decisions. Therefore, the goal of the 'Wise Policy Making' project is to develop methods and tools that can support policy making in dealing with the complex issues of steering towards sustainable societal wellbeing.

Wise decision making includes two main dimensions: ethics (concerning the ultimate goals we wish to pursue in order to live a fulfilling and joyful life) and intelligence (concerning the most effective way to get there). We define 'wise' policy making as the ability to effectively strive for sustainable societal wellbeing, hence requiring the combination of both ethics and ratio. In this report we will focus on how to augment collaborative decision making in order to tackle the complex issues involved in this ultimate goal of sustainable wellbeing. Research has already demonstrated specific tendencies and inclinations in human decision making, such as our irresolute and procrastinating response to long-term problems (Alexander and Brown, 2010) or our tendency in political discussions to be persuaded by arguments focusing on loss, rather than on gain (Loss aversion, Arceneaux, 2012). The branch of cognitive science dealing with these kinds of tendencies named them cognitive biases (Tversky and Kahneman, 1974; Kahneman, 2011). Cognitive biases are systematic inclinations or dispositions that affect decision making processes in ways that may make their outcomes ineffective. Biases in human (and animal) thinking and decision making are common in all people and contexts.

[1] As a working definition here we consider wisdom as being able (from an intellectual as well as an ethical perspective) to promote peoples' wellbeing, that is, to create a world in which people can flourish and be happy.

A large body of literature shows how biases affect thinking and decision making in a broad range of situations (e.g., Kahneman, 2011; Shafir and LeBoeuf, 2002).[2] We are quite blind to these (subconscious) inclinations of our minds (Pronin, Lin and Ross, 2002).

In line with their systematic and universal character many studies demonstrate that biases are also prominent in policy making (e.g., Arceneaux, 2012; Baron, 2009; Bellé, Cantarelli and Belardinelli, 2018; McDermott, 2004; Mercer, 2005; Shiller, 2015; Vis, 2011). For example, it has been demonstrated in many instances that policy makers tend to make risk-aversive decisions when they expect gains, whereas they accept taking more risk when facing losses (e.g., McDermott, 2004; Vis, 2011). These facts may count even more for the open/public part of the domain of policy making (Korteling, Sassen-Van Meer and Toet, 2020a). All in all such tendencies may lead to unbalanced and inaccurate collective decision making with ultimately suboptimal or wrong outcomes.

Research into mitigating these tendencies has explored different kinds of treatments or interventions (for an overview see e.g. Zenker and Dahlman, 2016; Larrick, 2004). Almost all of them that have so far been conducted, investigate the bias mitigating effect just after finishing the training interventions while using the same type of tasks in similar task domains that were also used during the training (Cheng and Wu, 2010; Clarkson et al., 2002; Larrick et al., 1990). However, to be truly effective, achieved effects of cognitive training should lead to *enduring* changes in the decision maker's choices (i.e. 'retention'). In addition people should be able to apply what they have learnt in situations and contexts beyond the specific training intervention in its (mostly narrow) context (i.e. 'transfer'). Therefore, in a previous phase of this ERP we carried out a systematic literature study of the available experiments on retention and transfer of bias mitigation interventions (Korteling, Gerritsma and Toet, 2021). We conclude that there is currently insufficient evidence that bias mitigation interventions will substantially help people to make better decisions under the practical conditions of real life. We have also carried out a literature study (Toet, Brouwer, van den Bosch, and Korteling, 2016) to gain insight into the extent to which individual characteristics (e.g. cognitive abilities, personality) affect a person's susceptibility to judgment and decision biases. The results of this study also show limited effects of these characteristics on cognitive biases. Finally, the experimental results so far have almost exclusively been obtained in restricted laboratory conditions. Here, simple tests are used that have specially been designed to measure the effects of bias mitigation interventions. Therefore, it is not clear how these results transfer to more complex real-life conditions.

The universal and robust character of biases may require the development of new approaches to improve human decision making lastingly and in a broad array of practical contexts. In the present report we try to provide a basis for such new approaches. This will be done for the context of modern deliberative democracy. This field is characterized by collaborative decision making aimed at joint support of as many involved parties or actors as possible. Here we focus on the solution of complex (political) issues with the involvement of multiple stakeholders (public,

[2] All biases mentioned are described in Appendix 1.

government, companies, interest groups, scientific- and domain experts etc.) who often have different and competing interests.

2 How to battle biases?

2.1 Introduction

The high number of biases involved in collaborative complex reasoning and decision making poses the question how to defend ourselves against these biases in a practical and feasible way. Korteling and Toet (2020, 2022) discussed several options or starting points on how to deal with biases in the decision making process. These are briefly presented below.

2.2 Bias mitigation approaches

2.2.1 Debiasing training (bias mitigation training interventions) focusing on specific biases:

Decision makers may be effectively debiased through training in specific domains. For example, experts can be trained to make very accurate decisions when decision making entails recognizing patterns and applying appropriate responses in domains such as firefighting, chess, and weather forecasting (e.g., Klein, 1998, 2008). However, there is barely any evidence to date of more general debiasing effects, across domains and different kinds of problems. That is, only a few studies have reported positive results with regard to the transfer (and retention) of bias mitigation interventions (Korteling, Gerritsma and Toet, 2021; Zenker and Dahlman, 2016). This means that, for instance, experienced weather forecasters may be able to predict snow with high accuracy, but they may show as many overconfidence flaws in their answers to trivia questions as other people. An exception to this rule is probably the mitigation of biases by education and training in statistics and probability calculation in fields that are heavily reliant on statistical data (Nisbett et al., 1987). As called for by Korteling et al. (2021), Larrick (2004), Ludolph and Schultz (2018), and Selier, Scopeletti and Morewedge (2019), more extensive studies on transfer effects are needed to investigate if interventions (of which gaming interventions are relatively promising) can beneficially aid decision making in a broad array of practical contexts.

2.2.2 Focus on mitigation of the main category of 'Evolutionary biases':

Most biases are structural and inherent to the working of the brain as a (biological) neural network (Korteling, Brouwer and Toet, 2018; Korteling and Toet, 2020, 2022). These 'neural' biases are inextricably linked to the basic system properties of our brain, without which our brain cannot function as a proper biological neural network. Because of this inherent, structural character, it is impossible to execute an order like 'Do not think of a pink elephant' or to control which information pops up in your mind when deliberating about an issue. In contrast, the other main category of cognitive biases, that is, the evolutionary biases (Self-interest, Herd thinking, Statistical blindness, Optimism) have a more 'functional' character. They may be conceived as a kind of inborn 'brain programmes' coding for certain preferences or inclinations that had value for survival and reproduction in primordial times. This means that, in contrast to the neural biases, the evolutionary biases are less fundamental (or hard-wired) with regard to the way neural networks work. This means that evolutionarily inherited survival tendencies, like striving for immediate reward (Hyperbolic discounting) and for self-interest at the expense of the community ('Tragedy of the commons'), may be suppressed and overcome by the

subject. This, however, will probably require substantial motivation, effort and perseverance (Korteling, Sassen-Van Meer and Toet, 2020a; Korteling et al., 2021)

2.2.3 Expertise development and domain knowledge:

Many authors, like Dane and Pratt (2007), Kahneman and Klein (2009) and Klein (1998, 2008) advocate that intuitive or 'heuristic' decision making can only be systematically successful when it is applied to problem areas in which the decision maker has a great deal of knowledge and experience. Outside those areas, the value of gut feeling is not much better than that of a guess. Experts therefore typically resist requests to make judgments about matters that fall outside their 'circle of competence'. They know that they don't know (Kahneman and Klein, 2009; Korteling, Sassen-Van Meer and Toet, 2020a). However, the development of skilled intuitions of an expert requires prolonged practice in a high-validity environment with feedback on actions that is both rapid and unequivocal (e.g., Kahneman and Klein, 2009). Deliberative democracy as well as most policy making takes place in a dynamic 'low-validity' environment with long time delays, dynamic circumstances and uncertain feedback. In addition, professional policymakers develop most expertise only with regard to the *formal* processes of policymaking. They do not develop much knowledge or expertise with regard to the *content* of the complex issues they have to deal with. More in-depth knowledge and expertise is hired if necessary or comes from lobbyists or stakeholders. Therefore, the building of sufficient expertise for adequate intuitive decision making on the complex topics of deliberative policy making is often not possible

2.2.4 Bias awareness training

Becoming aware of our limited cognitive capacities and our bias proneness, is likely a good first step focusing on creating a basis and willingness for further dealing with cognitive bias. Reducing the "Bias Blind Spot" (Pronin, Lin and Ross, 2002) may form a good (and necessary) basis for better or sound reasoning and effective decision making. However, bias-awareness is not *sufficient* for debiased reasoning and wise decision making. We have carried out several pilots in which the effectiveness of a quick bias awareness intervention could be evaluated.
This 'awareness' intervention consisted of a presentation on the nature and origin of biases combined with reflection and performing a few group-assignments. Several awareness workshops were held with professionals involved in military decision making. The workshop results, amongst others, showed that these interventions may result in improved bias awareness. However, discussions and evaluations afterwards indicated that this was certainly not the case for all participants with fair indications of the 'bias blind spot'. We now conjecture that a 'fairly high educational background' is a necessary - though certainly not a sufficient - condition that brief workshops lead to a noticeable enhanced degree of bias awareness (e.g., Korteling and Duistermaat, 2018; Vries, Boer-Visschedijk, Roijendijk, Veldhuis, and Rijk, 2021).

2.2.5 Prevention of the adverse effects of bias in decision making.

The notion that many biases arise from more or less hard-wired brain mechanisms, means that mitigating them by training interventions will always be an 'uphill-battle' requiring substantial motivation, effort and perseverance. Therefore we suppose that the most promising way of dealing with biases may be to improve the *environment* in which people make decisions instead of trying to directly augment the thinking capacities of people. For example, one could stimulate or impose the

use of certain working methods or aids with which the ingrained tendency to biased thinking can be *prevented or circumvented*. This means that instead of trying to directly improve people's ingrained thinking capacities the most effective approach of battling biases is likely based on support tools or working methods that prevent their adverse effects on decision making. So it concerns improvements of the *environment* or *context* of decision makers. This environmental support can be carried out from two different perspectives: a technical and a behavioural one.

In order to prevent biases, the project ERP Wise Policy Making focuses on both technological and behavioural forms of external support. Both kinds of environmental (or contextual) interventions may enhance the quality of the collaborative dialogue and reduce the chance of the intrusion of biases in decision making. A simple example of the *technical* approach are support tools ensuring that all crucial information is fairly documented, presented, analysed, and weighted. In this regard one could think of computer programmes that scrape the internet for specific information or (automated) checklists that help to gather all relevant information on a certain subject. An example of the second, *behavioural* or *cognitive* approach is working with 'premortems', that is, countering thinking tendencies such as optimism and overconfidence by actively imagining what could go wrong in a project (Klein, 2007). Another behavioural example is 'playing the devil's advocate' in order to counter confirmational thinking.

Both the technical and behavioural approaches of supporting collaborative decision making are developed in the *Wise Tank* work package of the ERP Wise Policy Making. The technical part of the Wise Tank work package (termed *Dialogue support*) aims at:

- A format that guides and structures a dialogue in such a way that critical facts and values are elicited in order to build a knowledge base for policy decision making.
- A digitalized representation of the content of the dialogue, such as rule statements, supporting and attacking arguments, values, definitions, and empirical facts, resulting in a digitalized network structure of well-arranged pros and cons of policy options.
- An analysis (through algorithms) of the represented content that distils the discrepancies and conflicts in the dialogue. In this way, the dialogue representation is dynamically analysed and updated so that it continuously directs the group to the discrepancies and conflicts in the dialogue that need to be addressed for a proper dialogue.

With regard to this latter possibility, it is generally supposed that for the execution of specific (narrow) cognitive tasks (logical, analytical, computational), modern artificial intelligence may be more effective and efficient than our biological intelligence (e.g. Korteling, Boer-Visschedijk, Blankendaal, Boonekamp and Eikelboom, 2021; Moravec 1988). Therefore, for the future, AI decision support tools may be developed that take 'weak' human components more out of the loop.[3] According to Haselager and Mecacci (2020), this may even concern ethical aspects of decision

[3] Ultimately, we expect that the development of digital decision support systems (augmented by artificial intelligence technology) may appear the most effective (and cost-efficient) way to improve complex human decision making

making. Since we appear to possess an underdeveloped capacity to act ethically and empathically, these authors conjecture that we do not need more intelligence but more ethics. Cognitive neuroscience and AI provide the knowledge and the tools to act more consistently upon ethical principles. The problem is not to formulate ethical rules, but rather to put them into practice. Cognitive neuroscience and AI are the technologies that may provide us with the 'moral crutches' to support people in their decision making regarding moral issues. Such technical tools could be applied by people engaged in teamwork or in a collaborative process, which certainly applies to the domain of political decision making. Behavioural knowledge is therefore an important basis for both the development of technical tools and the development of behavioural working methods. The present report explores the basics of a behavioural way to prevent biases and enhance wisdom in political decision making.

2.3 Conclusions

The notion that many biases arise from ingrained (structural and functional) brain mechanisms, means that mitigating them will often be very difficult. Therefore, the most promising way of dealing with biases is to improve the *environment or context* in which people make decisions, instead of trying to directly improve their ingrained thinking capacities. For example, one could use certain *technological aids* or *behavioural working methods* with which the ingrained tendency towards biased thinking can be *prevented, avoided or circumvented*. In the next two chapters we present the basics of the behavioural approach.

3 Approach 1: Effectively counteract multiple biases

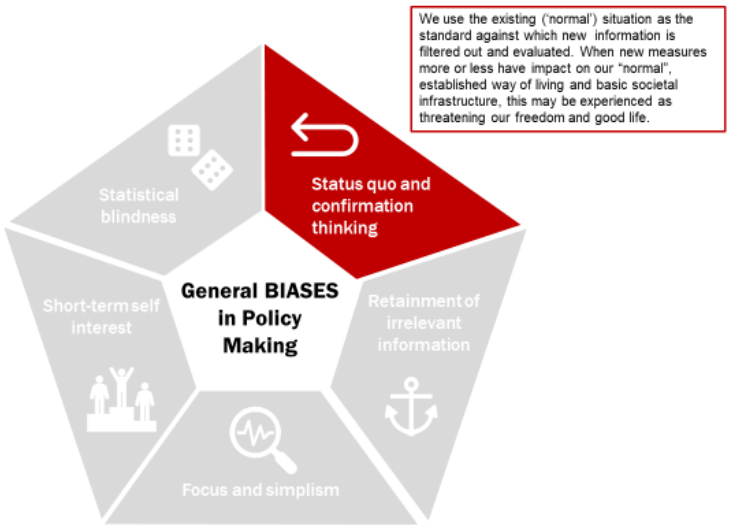
The scientific literature describes a multitude of cognitive biases, which are mostly phenomenologically described. Many cognitive biases resemble each other and discriminating the different types of biases is sometimes quite difficult. There is much overlap between the descriptions of different biases and one bias may be a more specific example of another, more broadly circumscribed bias. This means that whole groups or categories of similar biases, or 'different' biases that originate from some common underlying mechanism or principle, may be prevented by one kind of intervention in the decision making process. For example preventing biases related to our tendency to focus on limited information (e.g., Availability bias, Attention bias, Ego-centric bias, Experience bias, Focalism, Law of the instrument, Neomania, Salience effect[4]), may be countered by imposing working methods that aim at the inclusion of all (or as much as possible) relevant information.

Therefore, as a first step, this chapter proposes a first global description of five generic bias categories (groups/clusters of biases) in human judgement, reasoning and decision making. This clustering of biases into generic groups is based on our previous research and publications (Korteling et al., 2018; Korteling and Toet, 2020, 2022; Korteling, Sassen-Van Meer and Toet, 2020a, 2020b). In addition for each bias category some (global) working methods will be briefly suggested. These working methods will be elaborated further in a subsequent report, involving a handbook for process facilitators or group coaches how to organise and carry out wise collaborative decision making sessions.

3.1 Status-quo and confirmation thinking

This pertains to thinking from the starting point of the status quo or the existing situation. This existing situation involves what we already know or think, and what is experienced as normal. This 'normal situation' is used as the standard against which new information is 'filtered out' and evaluated. Typical biases that can be traced back to this tendency to think from the status quo and its confirmation are: Normalcy bias, Confirmation bias, Status quo bias, Default effect, Loss aversion, Endowment effect). When new measures have substantial impact on our 'normal', established way of living and basic societal infrastructure, this is experienced as a threat to our freedom and current life circumstances (Loss aversion). In contrast, people may like changes, as long as they are small (or gradual) such that they seem not to disrupt the overall status quo.

[4] See Appendix 1 for a glossary of all biases mentioned in the present report.



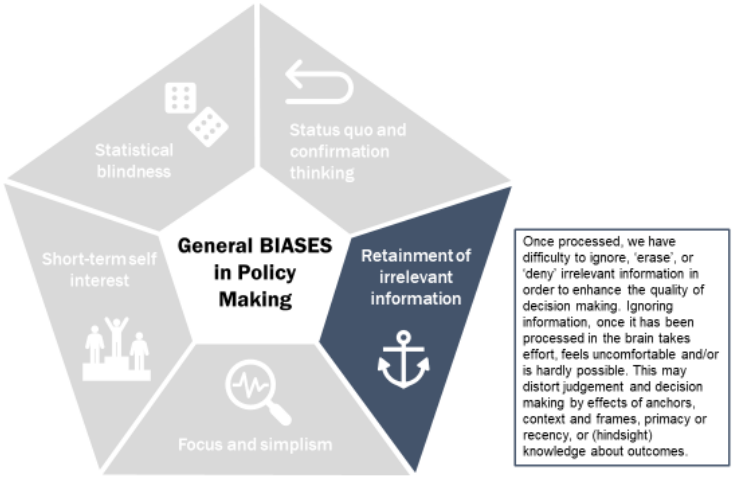
Practices dealing with 'Status quo and confirmation thinking'

Status quo and confirmation thinking leads people to think along conventional lines and creates a risk of missing new, or different approaches. Therefore the dialogue must not focus too much on the status quo rather than thinking outside of the box. This is especially so when new and creative solutions are required. These cases require an open mindset liberating decision makers from their usual thinking habits and standard beliefs and convictions. These are interventions that can help avoid confirmation thinking:

- Taking new and different perspectives. For example, taking the perspective of opponents, playing the devil's advocate to avoid tunnel vision (i.e. Confirmation bias).
- Make people aware of echo chambers (rabbit holes, filter bubbles) when they tend to ground their opinions or knowledge on information provided by social media.
- Create two groups: one generates ideas, and one evaluates them.
- Stimulate 'Critical thinking' (Glaser, 2017). Maybe also 'Lateral thinking' (de Bono, 1967) although some consider this pseudo-science (e.g. Melechi, 2020). Lateral thinking is used to break conventional thinking patterns and to broaden the scope of searches for new ideas: instead of going from A (problem) to B (a non-creative solution) using traditional step-by-step logic, one is forced to start somewhere other than A (lateral) and then go to creative solutions.
- Scholastic method: reverse the current/common assumptions or sentiments and try to honesty defend that reversed position, for example, 'why shouldn't we strive for less (instead of more) wisdom in policy making?'
- Look for disconfirming evidence. Conflicting evidence (this swan is black) provides more information than supporting evidence (I see only white swans). Subsequently, when disconfirming evidence is found, it is often challenging not to disregard this conflicting information (Semmelweis reflex). In order to promote disconfirmational thinking, let someone play the devil's advocate.

3.2 Retainment of (irrelevant) information

Once received, information is retained, or anchored, in the neural circuits of the brain (also termed: 'cognitive anchoring' or 'cognitive inertia'). A substantial proportion of this information may typically be irrelevant or counterproductive. In these cases the brain has difficulty ignoring an order to 'erase' or 'deny' irrelevant information. So, whatever has entered or was neurally activated cannot simply be (temporarily) discarded or deleted in order to enhance the quality of the decision making process. Ignoring information, once it has been processed in the brain takes effort, feels uncomfortable, and/or is hardly possible. This principle applies more as the information is dominant, lively, recent, or for whatever reason 'loaded' with affect. Retainment of irrelevant (aspects of) information may distort judgement and decision making by effects of anchors (Anchoring bias), context and frames (Framing), hindsight (Hindsight bias) or knowledge about outcomes (Outcome bias).



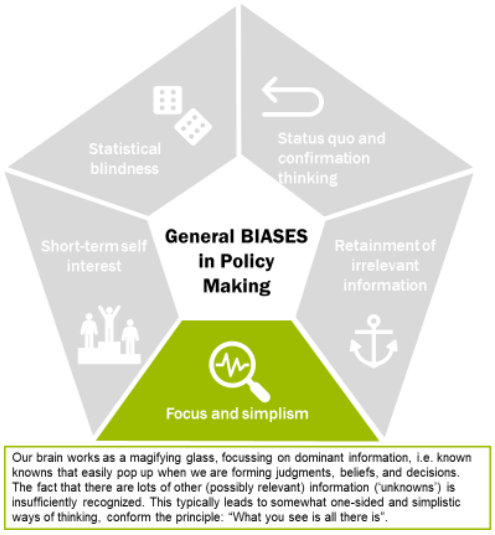
Practices dealing with 'Retainment'

Most dialogues or discussions are elaborations on what has already been said or contributed. This means that introducing new information (e.g. an example of something) or new lines of thought can be very defining for the future development of the dialogue. Care must therefore be taken to ensure that new information is introduced carefully and in a balanced manner. Ways to do this could be:

- Taking into consideration that all information brought into a debate has an order and our decisions tend to be biased towards the information that arrived first (the anchor). Therefore, one can make sure that various pieces of information in this respect are presented and processed in balance. This counts not only for the order of information, but also for contexts and frequencies.
- One should try to be careful with providing examples. When this is done it should be avoided that the dialogue turns to a focus on these examples.

3.3 Focusing and simplism

The brain works like a magnifying glass that focuses on dominant information. We focus on the known knowns that readily pop up in the formation of judgments, beliefs and decisions (e.g. Availability bias). The fact that there is lots of other (possibly relevant) information ('unknowns') is insufficiently recognized (i.e. the Knowledge illusion, a kind of 'blind spot' for ignorance in laymen). This typically leads to quick, somewhat one-sided, and simplistic ways of judgement and decision making, consistent with Kahneman's principle: "What You See Is All There Is" (WYSIATI). Focusing on known-knowns and simplism can aggravate our tendency to search for and find confirmation of the status-quo (see 1) and the Retainment (reinforcement) of irrelevant information (see 2).



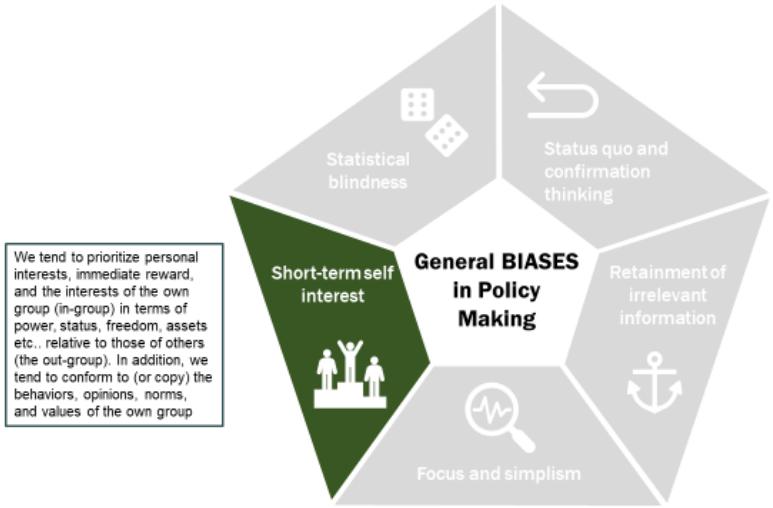
Practices dealing with 'Focus and Simplism'
People tend to underestimate the complexity of issues and challenges and tend to focus too much on what they themselves (coincidentally) know. They also tend to ignore system complexity (Meadows, 2009). It is therefore important that all available and relevant information is brought to the fore in a systematic and clear manner. Ways to do this:

- Infuse modesty into one's own knowledge in relation to the complexity of the subject.
- Gather and present all relevant information as equally and neutrally as possible.
- Use checklists to take into account all possible relevant factors in making decisions.
- Include alternative arguments: let's (also) talk about XYZ; instead of (only) about ABC.

3.4 Short-term self-interest and in-group conformity

Human behaviour is basically aimed at maintaining or strengthening the personal situation and conditions of living. Humans tend to prioritize personal interests and the interests of one's own group (in-group) in terms of power, status, freedom, assets et cetera, relative to those of others (out-group). We also have a preference for immediate reward compared to delayed reward, which makes it hard to withstand the temptation of direct gain (Hyperbolic time discounting).

Moreover, the sacrifices that have to be made in order to foster societal well-being are usually relatively high compared to individual immediate gains. Consequently, we tend to prioritize those actions that provide short-term individual benefit and postpone unpleasant but important actions for the collective (Tragedy of the commons). In addition, we are often more strongly influenced by the behaviours and opinions of the in-group than by our own personal views and attitudes.
One's own group determines acceptable norms and values and what is considered normal (In-group bias), and people tend to conform to these standards as a major reference for comparison (Conformity bias).



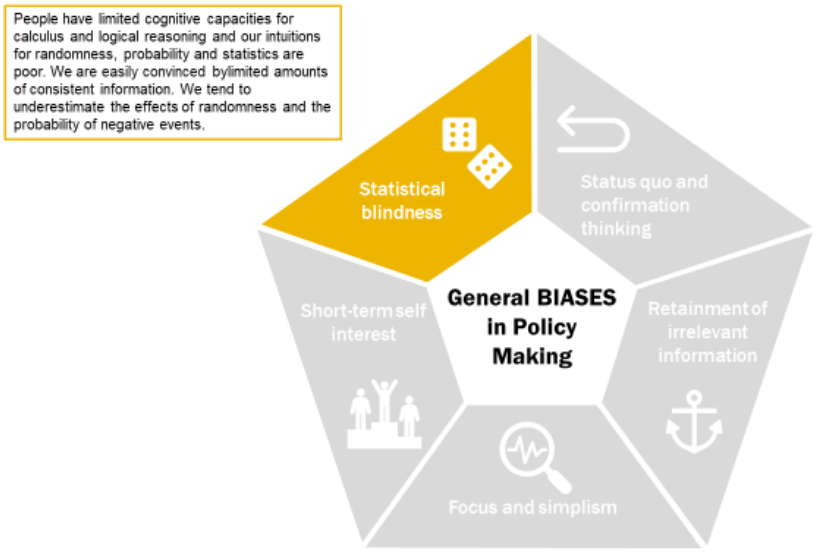
Practices dealing with 'Short-term self-interest and in-group conformity
People tend to put self-interest (or in-group interest) above the collective interest. That is why a lot of attention must be paid to creating a *mindset* in which achieving jointly shared goals and values is paramount. If this mindset is present in a substantial portion of the group, the tendency towards conformity with the in-group will lead non-conforming in-group members to align themselves with this dominant mindset. Therefore practices should be used to focus on a the common good. Ways to do this:

- Take the perspective of other people/society.
- Focus on what really leads to happiness (instead of falling into social comparison, copying other peoples' behaviour, or loss aversion).
- Prevent detrimental group pressure and copying behaviour and facilitate non-conformism and an open, critical attitude.
- Conduct a round of introductions in which one brings to the fore one's humanity, for example, one's dreams, one's favourite food or pastime – in order to mitigate one's tendency to focus on one's own role and perspective and have an interest in the perspective of the other.
- Employ Empathic Design (in which experts move towards users'/citizens' experiences) and Participatory Design (in which users/citizens move towards the experts' practices) (Steen, 2011).

- Use gaming and simulation of virtual reality as ‘empathy machines’.
  By immersing people in a virtual world you can give them realistic experiences that may be more effective than more abstract forms of information and education (Alsever, 2015; Korteling et al., 2021).
- Promote (some!) dissent and a willingness to change one’s own views and opinions. Do not suppress minority views.
- Take a longer perspective of time; imagine what a decision implies for your children. How would our ‘future-self’ assess our current decision (retrospectively)? (Börsch-Supan et al., 2018).
- Imagine or experience the future, for example, with the help of simulation, gaming and VR.
- Bring ‘far-away’ and ‘in-the-future’ effects into one’s here-and-now experience, for example, connect to people far-away (interview the child who works in the Bangladesh sweatshop) or simulate effects in the future (create an image of my family in 10 years’ time) (Intergenerational Foundation, 2018)

3.5 Statistical blindness

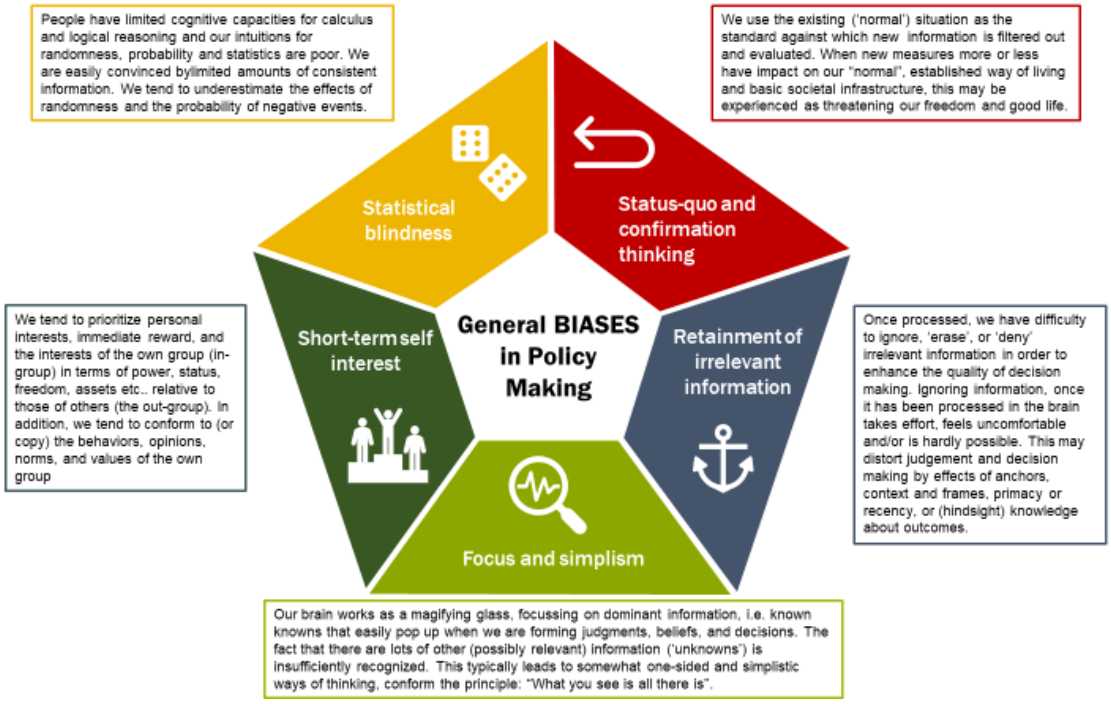
People have marginally evolved cognitive capacities for calculus and logical reasoning and our intuitions for randomness, probability and statistics are poor (Kahneman, 2011). This has resulted in various tendencies to draw erroneous conclusions on the basis of poor probabilistic and logical reasoning. The amount of cognitive information that our brain can consciously process (our ‘working memory’) is very limited. Most tasks involving probabilistic and logical reasoning require our full attention and we usually need a lot of time to execute them correctly and accurately (Korteling et al., 2021). Despite this fact, we have to draw inferences and build conclusions from complex, incomplete, or inconsistent (often numerical and probabilistic) data. We tend to overestimate the probability of positive or favourable outcomes (Optimism bias) and to underestimate the probability of negative events. ‘Black swans’ seldom factor into our planning, our economics, our politics, our business models and in our lives (Tail-risk blindness, Taleb, 2007). Instead of logically accounting for all these flaws by adjusting the confidence of our judgements downward, we overvalue the *consistency* of limited amounts of information, while being insensitive to possible random effects (Illusion of validity, Law of small numbers).



Practices dealing with 'Statistical blindness'

Statistic intuitions are generally poor, but they can be learned. Expectations or arguments based on quantitative and statistical intuitions must therefore always be critically tested against characteristics such as sample size, representativeness, long-term and tail-risks, and consistency. Ways to do this:

- Take care in presenting and explaining quantitative/statistical data because most people simply have a poor understanding of these kinds of data.
- Give sufficient attention to small odds that may disrupt a project, especially when these chances are cumulative. Small chances should also be taken more seriously than we usually do when the consequences are potentially huge (e.g., when they may lead to ruin).
- Take care not to draw conclusions too hastily, particularly when there is only a small amount of consistent data; consider the 'Law of small numbers'.
- Use models and provide appropriate statistical information, especially about the manifold possible ways that sufficient progress or failure of plans may be thwarted. Consider which possible setbacks may come your way.
- Counter optimistic tendencies by imagining what could go wrong in a project. ("premortems", Klein, 2007). Focus on all possible negative outcomes or unintended side effects of options.



3.6 Conclusions

The large number of biases involved in human reasoning and decision making makes it difficult to counter all of them. This begs the question how to defend against all these biases in a practical and feasible way. Since this large number contains many similar or overlapping biases, we suspect that entire groups or categories of similar biases can be efficiently countered by only one, or a limited number of interventions (working methods or forms) in the decision-making process. For instance, the behavioural technique 'playing the devil's advocate' may mitigate various biases that are related to the main group of biases called 'Status quo maintenance and confirmational thinking'. Therefore, we reduced the multitude of biases into five 'generic bias categories' that may affect collaborative decision-making processes. The result is a diagram of these bias categories. We posit that this structure of five generic bias categories may make it much more feasible to develop in the future a limited and well-arranged set of effective working methods to avoid (or circumvent) the emergence of bias in collaborative decision making. Finally, for each bias category we proposed some working methods that may be suitable for this purpose.

4 Approach 2: Wise Heuristics[5]

4.1 Introduction

Since biased thinking is inherent (neural, structural) and evolutionarily ingrained (functional) in the working of our brain (Korteling et al., 2018, 2020), it is a pervasive and difficult-to-mitigate phenomenon (Kahneman, 2011; Korteling, Gerritsma and Toet, 2021). Therefore, achieving rational and analytical (unbiased) thinking in cognitive tasks requires a high amount of support and effort. So, when solving complex problems, a possible way to improve decision making may be to focus on how 'normal' minds can still may make satisficing decisions when they have to work under time pressure and with limited amounts of knowledge and information. Such an approach is related to the notion of 'bounded rationality' (Simon, 1955). Humans are not endowed with an omniscient mind with endless computational power. Also the data we have with which to make a decision are often incomplete and uncertain. In practice, we usually only have limited thinking resources (brainpower, knowledge, memory capacity, time) and data (information, certainty). As a result, we cannot be expected to solve difficult problems, optimally incorporating all relevant aspects and data. It may be therefore considered useful or wise to adopt heuristics, wise principles, or educated guesses as a way to economise our limited cognitive faculties.

'Wise heuristics' is based on simple-but-rational, generic principles to produce good-enough solutions. It may be considered a pragmatic and flexible technique for decision making. This may be especially relevant in small-scale projects with limited resources or a tight time schedule. The 'wise heuristics' approach that is proposed here may be considered an alternative (or in some cases complementary) to working methods that strive for optimal results at all costs. It is simpler and more practical, since it does not focus primarily on the multitude of biases that have to be prevented. Instead, it starts with the deployment of powerful and wise principles. It does not aim to solve complex problems by pure logical reasoning, computing all probabilities and utilities of data, and explicitly taking into account biases. Wise heuristics are rational general principles that can be applied pragmatically and flexibly to a variety of complex problems or decision-making situations. We consider these 'rational heuristics' to be bias-resistant when used in the proper way.

This chapter must be seen as a first creative attempt to make an innovative step towards wise collaborative decision making, though it still needs further scientific substantiation. In the near future we plan to further develop and improve this 'wise heuristics' approach.

The selection of heuristics
The chosen generic heuristics, which are presented below, are based both on a limited literature study (e.g. Taleb, 2012; Engelfriet, 2014; Dobelli, 2011) and on an internal inquiry, supplemented with questionnaires and discussions with (project)

[5] The concept of 'heuristics' originates from the Greek 'heuriskein', meaning: the way to gain knowledge in a methodical way. This as opposed to the popular conception of heuristics in psychology as 'smart cognitive shortcuts' that help us to overcome our processing limitations and to deal with complexity or time constraints in decision making.

colleagues. An important criterion for selecting the suitable heuristics for this second approach was that they had to be simple and easily applicable. The chosen heuristics also had to be based on generally accepted efficiency principles or sensible starting points. So, they are not (necessarily) based on the intuitive shortcuts that normal (naïve) people may use in all kinds of practical situations (Gigerenzer & Todd, 1999; Gigerenzer & Gaissmaier, 2011; Gigerenzer & Goldstein, 1996; Simon, 1955). These latter kinds of 'fast and frugal heuristics' may result from our intuitive and natural inclinations, which are very sensitive to bias (Type 1 thinking, e.g. Kahneman, 2011). In contrast, the wise heuristics approach proposed here is based on rational principles that ultimately can be traced back to deliberate and analytical reasoning (Type 2 thinking).

A final criterion for a powerful and useful heuristic is the fact that the principle has already survived for a long period of time ('Lindy Principle').[6] According to Taleb (2012), the anticipated life expectancy of non-perishables (like ideas, technologies, concepts, behaviours) is proportional to the time it already has existed. These are the heuristics that have survived a long-enduring selection process. They may therefore be considered robust (or even 'anti-fragile') and are probably useful for many different domains. Many examples of non-perishables that do *not* pass the Lindy Principle can be found in hypes. Hypes are ideas, behaviours, concepts, or styles that become a fad or trend and spread quickly within a culture. These cultural trends or movements are impulsively and enthusiastically followed by many (groups of) people. They mostly last for a short period and seldom transfer to a remaining cultural or societal factor. So, most hypes do not stand the test of time and will almost certainly die out in a relatively short time.[7] Examples of non-perishables that pass the Lindy test are: gold, knives, dancing, stoic virtues, wine, cooperation, and the Bible. Examples of hypes are Steve Jobs schools, the presumed benefits of multi-tasking, and the many hypes in the area of food consumption, art, management styles, physical and mental fitness, wellness, stock markets and social media.

4.2 Examples of ubiquitously powerful heuristics:

In this section we present a first attempt at a list of selected heuristics that may enhance the quality of decision making in complex issues in a very simple and practical way. These heuristics can be introduced by the session leader of a collaborative decision-making dialogue when appropriate or relevant. In many cases these heuristics may also help avoid bias; sometimes they are even based on ways of thinking aimed at avoiding specific biases.[8] The heuristics mentioned are consistent with the literature on (practical) wisdom (e.g., Ferrari and Weststrate, 2013; Sternberg and Glück, 2019; Taleb, 2012). This means that they are defined by intelligent, rational, generic, logical, robust, and practical principles that can be effectively applied in the majority of complex issues.

[6] By the way, this Lindy principle is a heuristic principle in itself.
[7] For this reason it is misleading to regard so-called memes as the cultural analogues of genes.
[8] When heuristics are very closely related to dealing with specific biases (or categories) the relevant biases will also be mentioned.

It should be noted that the structured list of heuristic principles below is based on a 'top-of-mind inventory' of the authors and their colleagues, rather than a thorough review of the literature on this topic. Most of the heuristic principles inventoried are not new and have been previously formulated in other contexts and for purposes other than policy making. The list of heuristics below is clustered into seven (provisional) thematic categories: goals vs means, values vs facts, systems and complexity, test-of-time heuristics, less is more, social-psychological heuristics, heuristics to detect cognitive bias.

4.2.1 Goals vs means:

- We often tend to concentrate on intervening processes instead of on the final goals or results to be obtained. Therefore, we should take care to make a clear distinction between means (or measures) and goals (or ends). Next, we should avoid focusing too vigorously on the means and measures, instead always keep an eye on the intended final objective or desired result.
- Try to formulate and discuss the ultimate goals of the project as explicitly as possible.
- Start with a thorough inquiry into the question or problem to be solved. Is this indeed the core problem or is it possible to reformulate or redefine the problem in a way such that it can be solved more easily?
- Do not look for problems to match existing solutions. Rather, look for a solution that matches the problem you have.
- Be explicit about the time frame of each other's goals and perspectives and try to find a common denominator.
- Sometimes it may be useful to search for a common way that may serve several goals simultaneously.
- Set boundaries ('scope'): the boundaries of a system or goal determine who or what falls 'inside' or 'outside'. Boundaries are neither objective nor accidental. Boundaries are determined by ethical, political or pragmatic considerations.

4.2.2 Values vs facts

- Take care to distinguish between values and facts.
- Recognize that people have subjective perspectives determined by their values, goals, and roles. Knowing participants' situations and motivations allows you to understand them better.
- Try to make explicit and understand each other's personal values, goals and roles in relation to the overall collective problem and goal (of the project).
- Try to find (and emphasise) the common standards and similarities between the values and goals of the participants. Focus also on shared values and interests, for example, refrain from framing in terms of political left-right. The challenge is to facilitate a dialogue that will enable interlocutors to express their different values and listen to each other. This creates room for potentially shared values to emerge – even if opinions differ (Mersch, 2021).

4.2.3 Systems and complexity

- Take the complexity and interrelationships of things into account. The relationships and interaction between the parts of a complex system provide insight into the behaviour of the system as a whole.
- Single cause fallacy. Most problems originate from a combination of causal factors.
- Be modest about your knowledge and understanding of most topics, unless you are a very experienced expert. Limit the number of participants with little understanding of the problem and the domain (little substantive knowledge). That is, people who, due to their background and position, probably will not actively contribute to the concrete realisation of results and solutions.
- Context rules: human behaviour is largely determined by context and environmental cues and structures.
- Proportionality bias: impactful events are not necessarily explained by big causes. The lack of a clear single cause does not necessarily mean that the solution has to be complex.
- We have a poor imagination for things that do not (yet) exist or have happened ('End-of-history' bias). As a result, there is always much more change than you can imagine. This contributes to the fact that predictions of the future are usually wrong (i.e., 'prognosis illusion'). Take into account that our disregard of the unknowns leads to planning failures: projects usually take about twice (or more) as much effort (money, time, resources) than you think.
- By concentrating our attention, we tend to ignore other important (and noticeable) information. We should therefore be careful not to focus too much on details, contrasting information, or those aspects that vary or pop out.
- Distrust fluency or smoothness: things or solutions that are nice, smooth, clear and easy to understand are not necessarily better or truer.
- An 'x-or-y' problem can often be solved easily just by formulating it as an 'x and y' issue. When a problem is formulated as an 'x-or-y problem', x and y are a priori considered to be mutually exclusive, that is, when one is true, the other must be false. However, in many cases, x and y can both be true at the same time. For example, a discussion or decision problem may be analysed as: are these facts misinterpreted (x) or are they ignored (y)? Or: Is (unhealthy) candy too cheap or do people just find it very tasty? In many cases, both options can be true at the same time.
- Know your circle of competence and stick within it (Warren Buffet): If you are not an expert on the subject, assume that you probably only understand a small portion of the whole problem. Bring in arguments or information on subjects that fall within your circle of competence or ask an expert to do so.
- Do not stick to comparing one new option (A) with the status quo (which is often the continuation of an existing situation, or the default). Instead, encourage comparison with other possible, good alternatives, B, C, D, E etc.

4.2.4 Test-of-time heuristics

- Try to build further on the useful things that you already have and/or that already have proven their value. Do not be too hasty to start all over again (from scratch) but try to take advantage of what is already available to solve the problem.
- Take into consideration that replacing something (e.g. an old building) mostly destroys the value or usefulness of what is being replaced. In other words, don't throw the baby out with the bathwater.
- Use ideas and principles that have stood the test of time (Lindy principle). Modern ideas and hypes do not last long. Most immature or new ideas are fragile.
- Most innovations are likely to fail. So, do not be too hasty to rely on hypes and embryonal concepts and technologies (Neomania).
- Approach or deal with a problem from different perspectives. In doing so, try to look at the whole from a distance (taking distance), for example through the eyes of an outsider, by taking a more long-term perspective or a more global (instead of local) perspective. Viewing a problem from a somewhat greater distance, makes it easier to reach agreement on how to approach it.
- The burden of proof rests with those who change or intervene in the existing or natural course of events.

4.2.5 Less-is-more heuristics

- Simplicity is the ultimate form of sophistication! Unfortunately, we tend to think in terms of what should be *added*, instead of what should be removed. We overlook subtractive changes. (Adams et al., 2021). The 'Via Negativa' (a form of the Less Is More principle) counters this tendency by focusing on *subtractive transformation* (instead of additive transformation).[9] That is: think of solutions by what better *not* to do, instead of inventing more protocols, more control systems, more guidelines et cetera. Another heuristic following from the Via Negativa is to focus (first) on solving what is most needed, that is, the problems on the extremes instead of the more average ones.
- Use the principle of parsimony, or 'Occam's razor': the simplest explanation is usually the most likely one.[10]
- Information paralysis: be careful not to fall into the trap of endlessly searching/inventorying more and more information on a subject. The gathering of huge amounts of information may lead to obscuring the most essential or key information.
- Prevent 'Iatrogenesis': take unintended side-effects ('collateral damage') into consideration when intervening in a process. For example: take into consideration opaque heuristics, that is: routines or processes that seem of no value or to be redundant, but often may have lots of value just not in a clearly visible way. So, take care that solutions do not interfere with all kinds of possible opaque rationales. Organize a brainstorm session to identify the opaque advantages of the situation that is considered to need a change.

[9] Additive transformation refers to our tendency to solve problems by adding something.
[10] This is also termed 'Abductive reasoning'.

4.2.6 Social-psychological heuristics

- Give everyone an equal voice, also the ones with the most striking viewpoints.
- Invite and listen to minority views (Deep Democracy).
- Use the principle that good examples ensure good following.
- Motivation to achieve a goal becomes much stronger when the proponents of it have their 'skin-in-the-game'. This means that they have incurred personal risk (money, assets, danger or otherwise) by being involved.

4.2.7 Heuristics to detect cognitive bias

The probability of bias may be higher if people use a strong selection of facts, coloured choice of words, or a lot of tone of description, to convey a particular feeling or attitude. To help recognise this process, one can keep in mind the following indicators of (deliberately) influencing another's view in an unfair way (NJIT, 2020):

- Things are worded with the intent to oversimplify or over generalize. The author presents a limited view of the topic.
  - What facts has the author omitted?
  - What additional (often contextual) information is necessary?

- Words or frames are used to create positive or negative impressions (e.g. they claim that.. vs we know that….)?
  - What impression would I have if different words had been used?
  - The language of the document is often extreme; statements have an all or nothing connotations and frequently repeated.
  - The argument appeals more to emotions than to logic.

- How, and how well, a message, or argument, is (unconsciously) processed by the listener can be affected by many factors (Korteling, Duistermaat and Toet, 2018) :
  - Positive/negative association.
  - order of presentation (first and last are remembered better).
  - Style and language usage.
  - Repetition, which creates a feeling of familiarity that is not easily distinguished from truth (sometimes also called the 'mere exposure effect').

4.3 Conclusions

This chapter presented a first step towards the conception of a 'wise heuristics' approach to making wiser and less biased choices in complex collaborative decision making situations. This approach focused on powerful heuristic principles that can be used to easily reach satisfying results. The use of 'Wise heuristics' does not guarantee or aim at fully complete solutions. It provides a basis for improving the quality of complex decision making in a relatively quick manner. This way of quick decision making can be further optimised depending on time and budget (for example using a method based on counteracting biases, as covered in the previous chapter). In general, a wise heuristics approach (yet to be designed, operationalized, and developed) may be used in many different (parts of) decision making processes. For instance, it may be used on the level of a complex issue as a whole, but also for its sub-components.

5 Discussion and conclusions

Recent theoretical insights indicate that biases originate from very specific and ingrained (structural and functional) neuro-evolutionary working principles of our brain (e.g., Haselton, Nettle and Andrews, 2005; Korteling, Brouwer and Toet, 2018; Korteling and Toet, 2020; 2022; Tooby and Cosmides, 2005). Mitigation of these inherent thinking tendencies, for instance by training interventions or by building expertise will typically require much effort and time or may even be hardly possible. This problem is further aggravated by the multitude of different biases, all requiring mitigation by specific and dedicated forms of training or education. Therefore, we expect that the most promising way of dealing with biases in collaborative decision making processes (policy making dialogues) will be to improve the *environment or context* in which people make decisions. This instead of trying to directly improve people's ingrained thinking capacities. ERP Wise Policy Making explores two such contextual improvements. The first is the use of certain *technical tools* to improve decision making. The second focuses on the use of specific *behavioural (cognitive) working methods*. Both kinds of methods (technical, behavioural) aim to improve collaborative decision making through environmental interventions.

It is well-documented in the scientific literature that there is a large number of biases involved in human reasoning and decision making (e.g., Kahneman, 2011; Korteling and Toet, 2022). This poses the question how to defend against (or better: how to deal with) all these biases in a practical and feasible way. Since there are many similar or overlapping biases, we conjectured that whole groups or categories of similar biases may be prevented or avoided by only one kind, or a limited number of possible interventions. Therefore we proposed a global description of five generic bias categories (groups/clusters of biases) and put each of them in a diamond-diagram with simple icons. Working with this diagram of generic bias categories may make it more feasible to develop 'debiasing working methods' for collective dialogue and argumentation (with regard to policy making issues).

On the basis of this, we provided some first suggestions of feasible and adequate working methods for each bias category. These working methods can be used to circumvent or mitigate relevant cognitive inclinations. For each bias category, we identified several working methods that could be appropriate to oppose cognitive biases (to be elaborated in a subsequent report). This approach still has two drawbacks:

- The manifold possible working methods still require a great deal of psychological knowledge and skills from the process facilitator (coach) who guides the decision making dialogue. These processes often require quick interventions that are not always very simple or easy.
- In addition, some of these interventions usually require quite some effort and time to explain, execute and evaluate them.

These drawbacks of bias circumvention methods may at times make them inappropriate to deal with all kinds of policy making questions, situations or project phases. Therefore we provided a first attempt for a complementary and practical approach that fits well in small-scale projects that have limited resources at their disposal. This second approach is based on the use of 'wise heuristics'. Wise heuristics are defined as robust or generic principles of wisdom that have proven

their value through time. The wise heuristics approach can be used by facilitators who lead collaborative decision making sessions for policy making. This approach aims to use these principles to decrease the cognitive load of the participants involved in complex decision making, while at the same time help produce proper results. It may be considered a practical, efficient and flexible way to reach good decisions, particularly when working under time pressure or with lots of data. Wise heuristics do not guarantee fully correct or optimal solutions. As guiding principles for decision making, some wise heuristics may prove surprisingly robust across a wide array of problem types and problem contexts. In addition, wise heuristics can also be used to make an initial set of potentially suitable options or solutions. In a follow-up phase of a complex decision problem, more time-consuming, fine-grained and detailed analyses can then be performed to obtain a final choice.

An important criterion for selecting the presented heuristics was that they had to be simple and easily applicable. The chosen heuristics also had to be based on generally accepted (wise or 'rational') principles or sensible starting points. They are *reason-based*. So not on the 'shortcuts' (or rules of thumb) that ordinary people often use intuitively in daily life (take the best, take the most familiar, what pops up in your mind, follow the leader, be kind, etc…). This means that the 'Wise heuristics' approach should not be confused with notions related to 'bounded rationality' or 'fast and frugal heuristics' (e.g. Simon, 1955,1991; Gigerenzer and Goldstein, 1996; Gigerenzer and Todd, 1999. The usability of fast and frugal heuristics, is very sensitive to biased thinking (e.g. Kahneman, 2011). This is especially so in situations that do not match with the concrete primal circumstances within which our brain has developed (Korteling et al., 2018, Korteling and Toet, 2020, 2022). Hence, we expect that these intuitive "fast-and-frugal heuristics" will not be very beneficial for improving human decision making in complex and abstract policy making processes. The robust and generic heuristics that we have proposed have a fundamental rational and deliberate grounding. They are more sophisticated and generic and usually based on long rational traditions. These generic principles of wisdom for wicked problems come close to practical wisdom and lessons for a good life. They are based on ancient and classical lessons that have survived a long enduring selection process.

It is important to note that the heuristics presented in this report are not supposed to be a complete exhaustive set of heuristics compiled from an extensive theoretical study. So we have not listed all possible wise and heuristic principles that may be relevant to complex collaborative decision-making. The described heuristics therefore primarily form a first list of principles that can be used in the further development and testing of methods for Wise Policy Making. In a subsequent report, an initial set-up and results of pilots are planned in which we investigate how decision makers can most effectively deal with cognitive biases and benefit from wise heuristics.

6 Glossary

Anchoring bias: Biasing decisions toward previously acquired information. In this way, the early arrival of irrelevant information can seriously affect the outcome.

Attention bias (Attentional illusion, Gorilla-in-the-room effect). By concentrating our attention, we tend to ignore other important (and noticeable) information or (cf. Focalism).

Availability bias: the tendency to judge the frequency, importance, or likelihood of an event by the ease with which relevant instances come to mind.
(cf. Experience effect, Primacy- Recency- effect, Fluency-, Familiarity bias, Salience effect, Focusing illusion, Recency effect and Mere-exposure effect).

Bias blind spot: the tendency to recognize biased reasoning in others, while failing to notice one's own biases.

Cognitive dissonance: the tendency to search for and select consistent information in order to try to reduce discomfort when confronted with facts that contradict own choices, beliefs, and values.

Confirmation bias: the tendency to select, interpret, focus on and remember information in a way that confirms one's preconceptions, views, and expectations.

Conformity bias: the tendency to adjust one’s thinking and behaviour to that of a group standard

Default effect: the tendency to favour the option that would be obtained if the actor does nothing when given a choice between several options (cf. Fear for regret).

Ego-centric bias: the tendency to rely too heavily on one's own point of view and to fail to consider situations from other people's perspectives (cf. Ego-centric memory bias).

End of history bias: the tendency to underestimate future changes for your own life

Endowment effect: the tendency to value or prefer objects that are already owned over those that are not. (cf. Loss aversion).

Experience bias: the tendency to believe and remember things easier when they are experienced directly with our physical body and senses (or concrete pictures) instead of abstract representations (tables and statistics).

Focalism: the tendency to focus strongly on one dominant aspect of a situation. This can be an initial piece of information (cf. Anchoring bias, Primacy effect) or a focus on one striking aspect of the information (cf. Salience effect).

Framing bias: the tendency to base decisions on the way the information is presented (with positive or negative connotations), as opposed to just on the facts themselves.

Herd behaviour: the alignment of thinking and behaviour of individuals in a group without centralized direction

Hindsight bias: the tendency to erroneously perceive events as inevitable or more likely once they have occurred.

Hyperbolic (time) discounting: the tendency to prefer a smaller reward that arrives sooner over a larger reward that arrives later.

Illusion of validity: (Law of small numbers, Hot hand fallacy, Insensitivity to sample size, anecdotal evidence): the tendency of having faith in assessments if they are based on consistent information and data-agreement with little or no consideration of the factors that can limit their accuracy, like sample size.

In-group bias: the tendency to favour one's own group above that of others.

Knowledge illusion: the tendency in laymen to over-estimate own competence.

Law of small numbers: See Illusion of validity

Law of the instrument (Professional deformation, The-man-with-the-hammer effect): the tendency to overvalue a known tool or method and to ignore alternatives to it. "If your only tool is a hammer, then every problem is a nail".
Loss aversion: the tendency to prefer avoiding losses to acquiring equivalent gains.

Neomania: the tendency to overestimate the role or importance of new things (e.g., technology) for the future.

Normalcy bias: the tendency to underestimate both the likelihood of a disaster ('black swans') and its possible consequences, and to believe that things will always function the way they normally function (cf. Inductive reasoning error)**.**

Optimism bias: the tendency to overestimate the probability of positive events and underestimate the probability of negative events.

Outcome bias: the tendency to evaluate a decision based on its outcome rather than on what factors led to the decision.

Primacy effect: the first information that is received has more impact (a larger weight) than the information received later. This works especially if judgments or decisions must be made immediately.

Prognosis illusion: the tendency to overestimate our forecasting abilities when making predictions.

Proportionality bias: the tendency believe that big and influential things must have big causes.

Recency effect: the last information received has more impact (a higher weight) than previously received information

Salience effect: the tendency to pay much attention and attach much importance to salient or striking details, for example, when coming up with causes or explanations for an event (cf. Focalism).

Selective perception: The brain collects and interprets information selectively based on compatibility with its current state (i.e. views, knowledge, expectations).

Self-generation effect: the tendency to remember information that is self-generated relatively well.

Semmelweis reflex: tendency to reject new evidence or knowledge when it contradicts ones beliefs.

Single cause fallacy: the tendency to believe that cases or events have one single deeper reason or underlying cause.

Sleeper effect: the tendency to remember the information (or message) itself better than (the credibility) of the underlying source.

Status Quo bias: the tendency to maintain the current state of affairs.

Sunk-cost fallacy: the tendency to consistently continue a chosen course or investment with negative outcomes rather than alter it.

Survivorship bias: the tendency to focus on the elements that survived a selection process, while overlooking those that were eliminated.

Tail-risk blindness: the tendency to ignore possible rare events at the edges of a statistical distribution that often carry the greatest consequences, yet are also the most unpredictable. These “Black Swans” are big events that have never happened before and cannot be predicted, but still need our attention because the potential consequences may be huge.

Tragedy of the commons: the tendency to prioritize personal interests over the common good of the community

7 References

Adams, G.S., Converse, B.A., Hales, A.H., Klotz, L.E. (2021). People systematically overlook subtractive changes. *Nature* **592,** 258–261 (2021). https://doi.org/10.1038/s41586-021-03380-y

Adviescommissie burgerbetrokkenheid bij klimaatbeleid (2021). *Betrokken bij klimaat: Burgerfora aanbevolen*. Eindrapportage adviescommissie burgerbetrokkenheid bij klimaatbeleid.

Alexander, W. H. and Brown, J. W. (2010). Hyperbolically discounted temporal difference learning. *Neural Computation,* 22**,** 1511-1527.

Alsever, J. (2015). Is virtual reality the ultimate empathy machine? Retrieved 01-28-2021 from: https://www.wired.com/brandlab/2015/11/is-virtual-reality-the-ultimate-empathy-machine/

Arceneaux, K. (2012). Cognitive biases and the strength of political arguments. *American Journal of Political Science* 56(2). https://doi.org/10.1111/j.1540-5907.2011.00573.x

Baron, J. (2009). Cognitive biases in moral judgments that affect political behaviour, Synthese 172 (1), 7–35.

Bellé, N., Cantarelli, P. and Belardinelli, P. (2018). Prospect Theory Goes Public: Experimental Evidence on Cognitive Biases in Public Policy and Management Decisions. *Public Admin Review* 78(6). https://doi.org/10.1111/puar.12960.

Börsch-Supan, A.H., Bucher-Koenen, T. Hurd, M.D. and Rohwedder, S. (2018) Saving regret. Working Paper 25238. http://www.nber.org/papers/w25238.

Cheng F-F., Wu C-S. (2010). Debiasing the framing effect: The effect of warning and involvement. *Decision Support Systems* 49(3): 328-34. doi: 10.1016/j.dss.2010.04.002.

Clarkson P.M., Emby C, Watt VWS. Debiasing the Outcome Effect: The role of instructions in an audit litigation setting. AUDITING: A Journal of Practice. 2002;21(2):7-20. doi: 10.2308/aud.2002.21.2.7.

CPB-PBL-SCP (2021). Plan van-aanpak CPB-PBL-SCP Verankering Brede Welvaarts-denken in begrotingssystematiek van Kabinet en Kamer. Den Haag: Centraal Planbureau, Planbureau voor de Leefomgeving, Sociaal en Cultureel Planbureau . Brief Tweede Kamer: Plan van aanpak Kernset Indicatoren Brede Welvaart | CPB.nl.

Dane, E. and Pratt, M. G. (2007). Exploring intuition and its role in managerial decision making. *Academy of Management Review,* 32**,** 33-54.

De Bono, Edward. 2014. *Lateral Thinking: An Introduction*. London: Vermillion.
de la Torre-Ruiz, José M., Vera Ferrón-Vílchez, and Natalia Ortiz-de-Mandojana. 2014. "Team Decision Making and Individual Satisfaction With the Team." *Small Group Research* 45 (2):198-216. doi: 10.1177/1046496414525478.

Dobelli, R. (2011). *Die Kunst des klaren Denkens: 52 Denkfehler die sie besser anderen uberlassen.* Munchen: Karl Hanser Verlag.

Engelfriet, R (2014). Zo simpel kan het zijn. Zaltbommel: Haystack.

Ferrari, M., and Weststrate, N.M. (Eds) (2013). The scientific study of personal wisdom – From contemplative traditions to neuroscience. Dordrecht: Springer.

Flyvbjerg, B. (2009). Survival of the unfittest: Why the worst infrastructure gest built – and what can we do about is . *Oxford Review of Economic Policy* Vol. 25(3), 344-367.

Gigerenzer, G., and Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology,* 62**,** 451-482.

Gigerenzer, G. and Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review,* 103**,** 650-669.

Gigerenzer, G. and Todd, P. M. (1999). *Simple heuristics that make us smart,* New York, NY, USA, Oxford University Press.

Glaser, E.M.(2017). *Defining critical thinking*. The International Center for the Assessment of Higher Order Thinking (ICAT, US)/Critical Thinking Community. Retrieved 01-28-2021 from https://www.criticalthinking.org/pages/defining-critical-thinking/766.

Goldstein, D. G. and Gigerenzer, G. (2002). Models of ecological rationality: the recognition heuristic. *Psychological Review,* 109**,** 75-90.

Harris, S. (2010). *The moral landscape. How science can determine human values*. New York: Free Press.

Haselager, P. and Mecacci, G. (2020). Superethics Instead of Superintelligence: Know Thyself, and Apply Science Accordingly, AJOB Neuroscience, 11:2, 113-119, DOI: 10.1080/21507740.2020.1740353.

Haselton, M. G., Nettle, D., and Andrews, P. W. (2005). The evolution of cognitive bias. *In:* Buss, D. M. (ed.) *The handbook of evolutionary psychology.* Hoboken, NJ, USA: John Wiley and Sons Inc.

Intergenerational Foundation (2018). *How much dor retirees regret not saving more when they were younger – and what could it mean for today's young adults?* Retreived 01-28-2021 from: http://www.if.org.uk/2018/11/27/how-much-do-retirees-regret-not-saving-more-when-they-were-younger-and-what-could-it-mean-for-todays-young-adults/

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

Kahneman, D. and Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. *American Psychologist,* 64**,** 515-526.

Klein, G. (1998). *Sources of power: How people make decisions,* Cambridge, MA, USA, MIT Press.

Klein, G. (2007). Performing a project *pre*mortem. *Harvard business Review*. pp 18-19. Retrieved from: https://cltr.nl/wp-content/uploads/2020/11/Project-Pre-Mortem-HBR-Gary-Klein.pd.

Klein, G. (2008). Naturalistic decision making. *Human factors*, *50*(3), 456-460.

Korteling, J.E. van de Boer-Visschedijk, G.C., Blankendaal, R.A.M.., Boonekamp, R.C., and Eikelboom, A.R. (2021). Human- versus artificial-intelligence *Frontiers in Artificial Intelligence.* https://doi.org/10.3389/frai.2021.622364.

Korteling, J.E., Brouwer, A.M., and Toet, A. (2018). A neural network framework for cognitive bias. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2018.01561.

Korteling, J.E. and Duistermaat, M. (2018). *Psychological deception.*
Report TNO 2018 R11532. Soesterberg: TNO Defence, Safety and Security.

Korteling, J.E., Duistermaat M. and Toet. A. (2018). *Subconscious manipulation in hybrid/psychological warfare*. Report TNO 2018 R11543. Soesterberg: TNO Defence, Safety and Security.

Korteling, J.E., Gerritsma, J. and Toet, A (2021). Retention and Transfer of Cognitive Bias Mitigation Interventions: A Systematic Literature Study. *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2021.629354.

Korteling, J.E., Sassen-van Meer, J., and Toet, A. (2020a). *Neuro-evolutionary framework for cognitive biases*. Rapport TNO 2020 R10611. Soesterberg: TNO Defence, Safety and Security.

Korteling, J.E., Sassen-van Meer, J., and Toet, A. (2020b). *Neuro-evolutionary bias framework*. Report TNO 2020 R11451. Soesterberg: TNO Defence, Safety and Security.

Korteling, J.E. and Toet, A. (2020). Cognitive biases. In: *Reference Collection in Neuroscience and Biobehavioural Psychology.* Amsterdam-Edinburgh: Elsevier ScienceDirect. https://doi.org/10.1016/B978-0-12-809324-5.24105-9.

Korteling, J.E. and Toet, A. (2022). Cognitive biases. In S. Della Sala (Ed) *Encyclopedia of Behavioural Neuroscience*, *2nd* edition, ISBN 9780128216361, Elsevier. pp 610-619 pp 610-619 https://doi.org/10.1016/B978-0-12-809324-5.24105-9.

Larrick, R.P. (2004). Debiasing. In: Koehler D, Harvey N, editors. The Blackwell Handbook of Judgment and Decision Making; Oxford, UK: Blackwell Publishing; 2004. p. 316-37.

Larrick, R.P., Morgan, J.N., Nisbett, R.E. (1990). Teaching the use of cost-benefit reasoning in everyday life. *Psychological Science* 1(6): 362-70. doi: 10.1111/j.1467-9280.1990.tb00243.x.

Ludolph, R., Schulz, P.J.(2018). Debiasing health-related judgments and decision making: a systematic review. *Medical Decision Making.*38(1):.3-13. doi: 10.1177/0272989x17716672. PubMed PMID: 28649904.

McDermott, R. (2004) 'Prospect Theory in Political Science: Gains and Losses from the First Decade', *Political Psychology*, 25 (2), 289-312.

Meadows, D./H. (2009). Thinking is Systems – A Primer. London: Earthscan.

Melechi, A. (2020). Beware of lateral thinking. Retrieved 01-27-2021 from https://aeon.co/essays/lateral-thinking-is-classic-pseudoscience-derivative-and-untested.

Mercer, J. (2005) 'Prospect Theory and Political Science', *Annual Review of Political Science*, 8, 1-21.

Mersch, R. (2021). Wil je een goede ruzie? Heb het eerst over waarden, niet over feiten. *De Correspondent*, 24 juni. https://decorrespondent.nl/12497/wil-je-een-goede-ruzie-heb-het-eerst-over-waarden-niet-over-feiten/777940049568-cc9fa445.

Moravec, H. (1988). *Mind Children*. Harvard University Press.

Nisbett, R. E., Fong, G. T., Lehman, D. R., and Cheng, P. W. (1987). Teaching reasoning.. *Science*. 238 (4827): 625–631. https://doi.org/10.1126%2Fscience.3672116

NJIT (2021). *How to Evaluate Information Sources: Identify Bias. New Jersey Institute of Technology.* Retreived jan. 7 2021 from: https://researchguides.njit.edu/evaluate/bias

Oxford Reference (2021). Principle of parsimony. Retreived 01-27-2021 from: https://www.oxfordreference.com/view/10.1093/oi/authority.20110803100346221.

Pronin, E., Lin, D. Y. and Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin,* 28**,** 369-381.

Shafir, E. and LeBoeuf, R. A. (2002). Rationality. *Annual Review of Psychology,* 53**,** 491-517. doi: 10.1146/annurev.psych.53.100901.135213.

Shiller, R. (2015). *Irrational exuberance*. Princeton: Princeton University Press.

Sellier, A. L., Scopelliti, I. and Morewedge, C. K. (2019). Debiasing training transfers to improve decision making in the field. *Psychological Science*, 30(9), pp. 1371-1379. doi: 10.1177/0956797619861429.

Simon, H.A. (1955). Simon, H. A. (1955). A behavioral model of rational choice. *Q. J. Econ*. 69, 99–118. doi: 10.2307/1884852.

Simon, H.A. (1991). Bounded rationality and organizational learning. *Organization Science*, 2 (1), https://doi.org/10.1287/orsc.2.1.125.

Steen, M.(2011). Tensions in human-centred design, *CoDesign*, 7: 1, 45-60. doi: 10.1080/15710882.2011.563314.

Steen, M. (2013). Co-design as a process of joint inquiry and imagination. *Design Issues*, 29, 2 (2013), 16-28. https://doi.org/10.1162/DESI_a_00207.

Sternberg, R., and Glück, J. (Eds.). (2019). *The Cambridge Handbook of Wisdom* (Cambridge Handbooks in Psychology). Cambridge: Cambridge University Press. doi:10.1017/9781108568272.

Taleb, N.N. (2007). *The Black Swan: The Impact of the Highly Improbable.* Penguin*. ISBN* 978-0-1410-3459-1.

Taleb, N.N. (2012). Antifragile – Things that gain from disorder. New York: Random House.

Todd, P. M., and Gigerenzer, G. (2000). Précis of Simple Heuristics That Make Us Smart. *Behavioral and Brain Sciences,* 23*,* 727–780.

Toet A, Brouwer AM, van den Bosch, K, Korteling JE. (2016). Effects of personal characteristics on susceptibility to decision bias: a literature study. *International Journal of Humanities and Social Sciences* 8(5), 1-17.

Tooby, J. and Cosmides, L. (2005). Conceptual foundations of evolutionary psychology. *In:* Buss, D. M. (ed.) *Handbook of evolutionary psychology.* Hoboken, New Jersey: John Wiley and Sons, Inc.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science,* 185, 1124-1131.

Vis, B. (2011). *Prospect theory and political decision making*. *Political Studies review*: 2011 (9), 334–343. doi: 10.1111/j.1478-9302.2011.00238.x

Vries, D. de, Boer-Visschedijk, G.C., Roijendijk, L., Veldhuis, G.; Rijk., R. van (2021). *Effectiviteitsstudies interventies Adaptieve begeleider.* Draft Report TNO Soesterberg: TNO Defence, Safety and Security.

Zenker, F., and Dahlman, C. (2016). Reliable debiasing techniques in legal contexts? Weak signals from a darker corner of the social science universe. In F. Paglieri, L. Bonelli, and S. Felletti (eds.) *The Psychology of Argument: Cognitive Approaches to Argumentation and Persuasion*, London: College Publications.173–196.