Preference stability over time with multiple elicitation methods to support wastewater infrastructure decision-making

Lienert, Judit*; Duygan, Mert; Zheng, Jun.

Swiss Federal Institute of Aquatic Science and Technology, Eawag, P.O. Box 611, CH-8600 Duebendorf, Switzerland

* Corresponding author: judit.lienert@eawag.ch; Phone: +41 58 765 5574

Co-authors: mert.duygan@usys.ethz.ch¹; zhengjun516@hotmail.com¹

Abstract

We used a multi-method and repeated elicitation approach across different stakeholder groups to explore possible differences in the outcome of an environmental decision. We compared different preference elicitation procedures based on Multi Criteria Decision Analysis (MCDA) over time for a water infrastructure decision in Switzerland. We implemented the SWING and SMART/SWING weight elicitation methods and also compared results with earlier stakeholder interviews. In all procedures, the weights for environmental protection and well-functioning (waste-)water systems were higher than for cost reduction. The SMART/SWING variant produced statistically significantly different weights than SWING. Weights changed over time with both elicitation methods. Weights were more stable with the SWING method, which was also perceived as slightly more difficult than the SMART/SWING variant. We checked whether the difference in weights produced by the two elicitation methods and the difference in their stability affects the ranking of six alternatives. Overall an unconventional decentralized alternative ranked first or second in 92% of all elicitation procedures, which were the online surveys or interviews. For practical decision-making, using multiple methods across different stakeholder groups and repeating elicitation can increase our

¹ Former Eawag employees

confidence that the results reflect the true opinions of the decision makers and stakeholders.

Keywords

Behavioral OR; weight elicitation; multiple criteria analysis; online survey; OR in environment and climate change; stakeholder preferences

1 Introduction

It is often crucial to choose decision-making processes that reflect values well. Environmental policy decisions are especially challenging and are increasingly addressed with Multi Criteria Decision Analysis (MCDA); for recent reviews see e.g. Gregory et al. (2012); Hajkowicz & Collins (2007); Huang et al. (2011); Mendoza & Martins (2006). Here, we focus on water infrastructures for which we devoted considerable efforts to include stakeholder preferences (Lienert et al., 2013, 2015; Scholten et al., 2015; Zheng et al., *subm.*). Water and sewer networks or wastewater treatment plants provide many services, including public health and protection from urban flooding or environmental pollution. These infrastructures are long-lived and aging; engineering planners face “daunting” future uncertainties (Milly et al., 2008). In many OECD countries, annual investment needs for water infrastructures are estimated at 0.75% of GDP (Cashman & Ashley, 2008), i.e. US\$ 300,000 million annually (OECD, 2012). A quarter of Swiss sewers need renovation, and the depreciation of public water infrastructures in the next 40 years amounts to CHF 81 billion (Maurer & Lienert, 2014). So it is almost inevitable that Swiss society will have to make major investments. MCDA relies on eliciting preferences from stakeholders² (e.g. Belton & Stewart, 2003; Eisenführ et al., 2010; Keeney & Raiffa, 1976). Environmental decisions affect many people, are usually financed by public money, are often long-term and may thus include the interests of future generations, and cannot easily be reversed. There are often very different views about the decision, resulting in conflicts of interest (Gregory et al., 2012). It thus seems especially important for environmental decisions that the elicited parameters reflect the preferences of different stakeholders and the public and that

² For the sake of conciseness, we use “stakeholder” synonymously to “decision maker”

these are stable over time and persist beyond a single situation (Gregory et al., 2012, p. 210), especially if they affect long time ranges. However, empirical evidence suggests that human beings easily become biased during preference elicitation (for reviews see, for example, Hämäläinen et al., 2013; Montibeller & von Winterfeldt, 2015; Morton & Fasolo, 2009; Payne et al., 1992; Slovic, 1995). The behavioral aspects of environmental decision-making have hitherto received comparatively less attention (Hajkowicz, 2012; Hämäläinen, 2015). For example, strategic bias (for references see Hämäläinen, 2015) is a major concern in public policy and environmental choices where real or perceived conflicts of interest often exist, but this has only recently been tested in an environmental context (Hajkowicz, 2012). To overcome strategic bias, it would seem important to include the views of different kinds of stakeholders, including the broader public (Gregory et al., 2012). Another specific example is groupthink (Janis, 1972), which can be an important bias in environmental modeling and decision-making, but the literature on this is very limited (Hämäläinen, 2015). A recent textbook about environmental management choices mentions important biases such as representativeness, availability, sunk costs, anchoring, overconfidence and motivation (Gregory et al., 2012, p. 29–31). The main message about how to deal with biases is to be aware of them, to use multiple framings to guard against affective responses and to use multiple elicitation methods (Gregory et al., 2012, p. 212). Additionally, we suggest that further research is required. The MCDA behavioral literature also lacks work on preference stability over time. For instance, we found no reference to preference stability in recent reviews of behavioral operations research (BOR; Hämäläinen et al., 2013; Montibeller & von Winterfeldt, 2015).

Against this background, the main aim of this paper is to explore the extent to which a multi-method and repeated elicitation approach that includes different stakeholders can contribute to increasing the confidence in the results of a decision-making process.

Using the example of Swiss water infrastructures, we investigate whether the results of an earlier MCDA are supported by a larger public prepared to finance these investments, independently of the elicitation method and a time gap of one month. In an earlier

project, namely SWIP (“Sustainable Water Infrastructure Planning”³), we interviewed twenty stakeholders using the popular SWING method (e.g. Eisenführ et al., 2010; Scholten et al., 2015), or a variant of SMART/SWING (Mustajoki et al., 2005; Zheng et al., subm.); the latter was perceived as less difficult, supporting the claim of Mustajoki et al. (2005). In the two interview series, we found patterns for the weights of objectives and we could identify viable wastewater and water supply alternatives, despite a very large uncertainty of predictions, stakeholder preferences and four future socioeconomic scenarios (Scholten et al., 2015; Zheng et al., subm.). In this paper, we build on this earlier experience and address the following research questions:

- Are preferences stable over time?
- Do SWING and the SMART/SWING variant differ in the elicited weights and in preference stability?
- What are possible explanations for the observed weight profiles or preference instability when elicitation is repeated? As examples, are demographic variables or environmentally friendly behavior possible explanatory factors? Do experts who have more knowledge or experience in the domain of wastewater have different and/or more stable preferences than lay people? Do preferences change if new knowledge is obtained between the first and second survey? Is preference stability related to the perceived difficulty of the questionnaire?
- If preferences do differ, does this affect the ranking of alternatives in an MCDA?
- Finally, if substantial public decisions are made, agreement between stakeholders and the public is important. In our water infrastructure case, do preferences about the importance of objectives from earlier stakeholder interviews agree with the preferences of the general public and of experts in the water field?

To tackle these questions, we conducted a public online survey in an experimental set-up with a simplified objectives hierarchy. Online tools overcome the problems of personal interviews by allowing access to the larger sample needed for the experimental research proposed here. They also speed up elicitation and allow involving the public. On the other hand, their validity and biases have been criticized (e.g. Marttunen &

³ <http://www.eawag.ch/en/departments/sww/projects/sustainable-water-infrastructure-planning-swip/>

Hämäläinen, 2008). The use of online tools thus requires careful design and cross-validation with results from other elicitation methods.

The remaining paper is organized as follows: In Section 2 we review the literature on preference stability. Our research approach and statistical tests are presented in Section 3. The results are presented in Section 4 and discussed in Section 5. We summarize our main conclusions in Section 6.

2 Literature review

2.1 Validity of different weight elicitation methods

Preference elicitation research has focused strongly on weights, e.g. the splitting bias and range insensitivity (Hämäläinen & Alaja, 2008; Jacobi & Hobbs, 2007; Montibeller & von Winterfeldt, 2015; Morton & Fasolo, 2009; Weber & Borchering, 1993). For instance, there is compelling evidence that different elicitation methods can lead to divergent results (Borchering et al., 1991; Bottomley & Doyle, 2001; Bottomley et al., 2000; Mustajoki et al., 2005; Pöyhönen & Hämäläinen, 2001; van Ittersum et al., 2007). To illustrate this, AHP is known to cause a larger spread of weights compared to other methods (Pöyhönen & Hämäläinen, 2001; and references therein), and AHP weights can be unevenly dispersed over the weight range (Belton, 1986). Whilst the impact of elicitation methods on weights is undisputed, there does not yet seem to be any consensus about the most valid method.

On the basis of a review, van Ittersum et al. (2007; Tab. 1) explain the apparent lack of validity of elicitation methods by setting up the proposition that different methods measure different dimensions of the importance of an attribute or objective, namely its (a) *salience*, (b) *relevance*, and (c) *determinance*. (a) Methods such as free elicitation, which are not common in MCDA, do not include information about objectives and thus measure only their *salience*. (b) Methods such as direct rating, direct ranking, point allocation, and AHP contain specific information about objectives, excluding their level. They measure the *relevance* of the objectives on the basis of the values and desires of the decision-makers. (c) Additional information about the level and ranges of each objective helps decision-makers to determine the importance of an objective in addition

to personal values. Trade-off, SWING, and conjoint methods are forms of elicitation that measure *determinance*. The proposition of van Ittersum et al. (2007) is strongly supported by their meta-analysis (see Tab. 2) as well as by an Internet survey that compared direct point allocation, SWING, trade-off, AHP, and SMART weight elicitation (Pöyhönen & Hämäläinen, 2001). This study showed that the weights elicited with the first three methods belonging to group (c) were similar, while SMART and AHP from group (b) tended to generate divergent weights. Despite compelling evidence for the multi-dimensionality of attribute importance, this concept does not explain all the inconsistencies between these methods (see Tab. 3; van Ittersum et al., 2007). The problem of possible inconsistencies between methods measuring the same dimension therefore remains unresolved.

2.2 Preference stability

As for any method, it is desirable that MCDA methods are reliable. One characterization of “good elicitation methods” could be that they produce stable weights – across methods and over time. On the other hand, it seems that preferences can easily change due to various influencing factors, and a large body of psychological studies indicates that preferences are affected by the elicitation method and its framing. Thus, some decision analysts postulate that preferences are constructed during elicitation (e.g. Belton & Stewart, 2003; Gregory et al., 1993; Hoeffler & Ariely, 1999; Payne et al., 1992; Slovic, 1995). In contrast, some economists presume that existing preferences can be captured by means of willingness to pay (WTP) surveys, for example.

Economic approaches have a lot in common with MCDA: in environmental decisions, both aim to derive valuations from respondents for an environmental good such as clean air or water quality. Economists model preferences by means of a random utility model (McFadden, 1974). As in MCDA, attribute-based economic methods such as choice experiments describe a decision situation as a function of its attributes (Brouwer et al., 2010). For a critical comparison of MCDA with economic valuation, see Gregory et al. (1993). For both schools, it is important to disentangle the various factors that may influence preference stability. Our own review suggests a striking lack of MCDA behavioral research on preference stability over time; this is also indicated by its complete absence in recent BOR reviews (Hämäläinen et al., 2013; Montibeller & von

Winterfeldt, 2015). Therefore, we also draw on the economics literature on preference stability. A summary of our literature review is given in the Supplementary Material (Section SI-3.1).

Economists have carried out various experiments that address reliability: “*Reliability* deals with whether a survey item measures something other than random noise” (Jorgensen et al., 2004; also see Liebe et al., 2012; and references therein). To investigate the reliability of contingent valuation (CV) surveys and choice experiments (CE), the test retest method is often used. It measures correlations between, for instance, responses to CV questions at two or more different points in time, or correlations between the mean WTP of the surveyed population. In general, CV and CE surveys applied to environmental valuations, consumer research or healthcare indicate a fair to substantial test retest reliability (see review sections by Jorgensen et al., 2004; Liebe et al., 2012). For instance, preferences about river restoration in Austria (Bliem et al., 2012) or water scarcity in Australia (Brouwer et al., 2010) did not change after one year. However, preferences can also be unstable: thus the WTP for natural areas in Flanders changed between two surveys one year apart (Schaafsma et al., 2014). There is some evidence that the parameter estimates of Expected Utility Theory (EUT; i.e. risk attitudes) are stable over time (reviewed in Zeisberger et al., 2012), but a comparison of methods seems to produce large biases and noise (Hey et al., 2009). However, our work in this paper does not focus on uncertainty and EUT. Rather, we are interested in determining possible reasons for preference changes under conditions of certainty.

2.3 Demographic and other personal factors

Preferences can be influenced by personal and demographic factors such as gender, age, income or environmental attitude. Gender, as an example, was a good predictor to explain preference stability between two CE surveys of out-of-hours healthcare for children (San Miguel et al., 2002). However, preference stability over time does not usually seem to be correlated with socio-demographic factors. For instance, in two CE surveys of environmental amenities, preference changes after one year were not significantly correlated with the financial situation of the respondents and other socio-demographic variables (Liebe et al., 2012; Schaafsma et al., 2014). Similarly, although

some studies did find a significant correlation between the WTP and explanatory variables, namely gender, income or environmental interest (e.g. Bliem et al., 2012; Brouwer, 2006), these differences were not related to preference stability over time.

2.4 Past experience, existing knowledge and expert knowledge

The existing knowledge or past experience of respondents could affect preference stability. Weight elicitation experiments with students indicate that knowing more about buying a car, in this example, increases preference stability over time (Bottomley & Doyle, 2001; Bottomley et al., 2000). A possible explanation is that experts are liable to hold stronger opinions, which are more stable, whereas novices have to construct their preferences during the experiment. Although this construction process may lead to the observed response variation in repeated experiments (Hoeffler & Ariely, 1999; San Miguel et al., 2002), this hypothesis was not confirmed in real world experiments. For instance, previous experience with out-of-hours healthcare before the experiment did not affect preference stability about healthcare management plans when the survey was repeated after two months (San Miguel et al., 2002). Likewise, in a CE about health risks related to poor bathing-water quality in Dutch rivers, preference stability did not differ if respondents had more knowledge about existing quality standards for bathing water (European blue flag system), or if they had more experience than others, such as by swimming more often in open water or becoming ill after swimming (Brouwer, 2006).

2.5 Learning, new experience and external events

There is fairly good evidence to suggest learning may affect preference formation. For instance, an in-depth experiment over several months concluded that the subjects' preferences are influenced by learning processes (Hanley et al., 2006). Repeated choice experiments (CE) about purchasing decisions indicate that consumers construct their choices when entering a new domain, while preferences stabilize with growing experience (Hoeffler & Ariely, 1999). A population survey in Australia also concluded that learning occurred: respondents were increasingly more certain about their WTP to reduce water scarcity problems in five repeated CE (Brouwer et al., 2010). However, the initial hypothesis that preferences would be refined due to this learning effect had to be rejected, since preferences remained stable across the experiment. A small learning

effect was also found in a CE about onshore wind power in Germany (Liebe et al., 2012). In this case, preferences were only moderately stable, with 59% of the choices being identical in the second survey after 11 months.

Over a time span of several months, one might expect external events to affect preferences, but this hypothesis finds scarce support in the literature. The financial crisis did not lead to any changes in the respondents' WTP for onshore wind-power generation (Liebe et al., 2012). Similarly, preferences concerning the WTP to reduce health risks associated with poor bathing water quality were stable despite a very unusual external event (Brouwer, 2006): Here, an earlier CE was repeated after six months, and by chance this coincided with the hot summer of 2003, when extremely high water temperatures had caused the closure of bathing sites in Dutch rivers.

2.6 Strength of preferences

People who have strong preferences might be expected to have more stable preferences than those whose weights are more evenly distributed. We found only one experiment in the literature concerning out-of-hours healthcare which confirmed this idea (San Miguel et al., 2002). One explanation could be that people with a more extreme weight distribution may see the respective objectives as more salient, thus increasing the stability of their weight profiles (based on Hoeffler & Ariely, 1999). This hypothesis could be important for MCDA because there seems to be a tendency for people to distribute equal weights; this phenomenon is also known as an “equalizing bias” (Montibeller & von Winterfeldt, 2015).

2.7 Perceived difficulty of method

Another reason for preference stability to differ between methods could be related to their difficulty. An experiment on consumer decisions revealed that preferences were more stable if choices were hard than if they were easy (Hoeffler & Ariely, 1999). The authors postulate that difficult and repeated trade-offs might help people to consolidate their preferences during their decision-making process, thus making them more stable. We found only two other studies that mention the difficulty of the elicitation process, but they did not discuss it: preference stability concerning out-of hours healthcare after two months was significantly correlated with the perceived difficulty of a CE questionnaire

(San Miguel et al., 2002). Similarly, a CV experiment concerning the bathing-water quality of rivers showed a significant negative correlation between increasing perceived difficulty of the questionnaire and preference stability (Brouwer, 2006).

2.8 Summary of review

Our review reveals a lack of MCDA behavioral research about preference stability, although some lessons can be learnt from the economics literature. Clearly, different elicitation methods can produce different weights, possibly because they measure different dimensions of the importance of an objective (van Ittersum et al., 2007). Correlations between preferences from the first and second surveys were often 'moderate' (linguistic classification of coefficients in Liebe et al., 2012): the strength of agreement between the two tests was usually >0.6 (Bottomley & Doyle, 2001; San Miguel et al., 2002), and >0.5 in 14 of 17 reviewed papers (Jorgensen et al., 2004; see Section SI-3.1). The observed preference changes could be related to socio-demographic factors, past experience, knowledge and expertise about a topic, learning, the occurrence of external events, the strength of preferences and the difficulty of the elicitation method. However, the literature on these topics is scant and the evidence even contradictory.

Finally, we wish to point out that even if individuals do change their preferences over time, this might not affect aggregated samples and the results of a population survey. This factor is discussed in the literature, which usually showed no significant differences between aggregated sample averages, despite large individual variability (Bottomley & Doyle, 2013; Jorgensen et al., 2004; Liebe et al., 2012; Zeisberger et al., 2012). A main aim of our research in this paper is to determine the preference stability for two typical MCDA weight elicitation procedures and find possible explanations for the observed weight profiles or preference instability when elicitation is repeated (see research questions in Section 1). In this way, we hope to contribute to closing some knowledge gaps in Behavioral Operational Research and to point out directions of future research.

3 Material and Methods

3.1 Design of online survey and sample recruitment

We used two questionnaires, each in English and German, one for SWING and one for the SMART/SWING variant (Section 3.2 (B)). A market research company (www.respondi.com.de) recruited respondents from the German-speaking part of Switzerland. They belong to a pool of people that have an agreement with the company to answer surveys on a regular basis. An e-mail invitation was sent to 708 people in July 2013 and 4–5 reminders followed within one week. All 807 Eawag⁴ employees were also invited by e-mail with a reminder after one week. Respondents were randomly assigned to one of the two methods. One month later, those who had fully completed the first questionnaire received a second one with the same elicitation method.

3.2 Questionnaires

The first questionnaire consisted of four sections.

(A) Knowledge and experience about wastewater designed to determine whether experts have different and/or more stable preferences than lay people: to determine their knowledge, we asked ten multiple choice questions, one for each sub-objective (Fig. 1), based on a survey (Veronesi et al., 2014). For instance: “after how many years do sewer pipes need to be replaced?” or “what type of disease is typically related to direct contact with wastewater?” (full list see Supplementary Material; Tab. SI-1). Each fully correct answer received one point. We asked respondents to specify their experience with seven wastewater topics, either as their own direct experience (two points), or information from mass media/ other sources (e.g. flooding of cellars with wastewater; Tab. SI-2). The highest score was $3 \times 7 = 21$.

(B) Main section to elicit preferences: we used a simplified objectives hierarchy from the SWIP project mentioned earlier (Lienert et al., 2015). Elicitation started from the bottom-up by comparing five pairs of sub-objectives followed by five main objectives (Fig. 1). The participants thus understood these after having considered the sub-objectives. To avoid a splitting bias, we used two sub-objectives for each main objective and kept

⁴ Swiss Federal Institute of Aquatic Science and Technology

the same design and text length (Schuwirth et al., 2012; Box SI-1–8). We described the best and worst possible case of each sub-objective and the status quo (Fig. 1; Tab. SI-3). We used data from a real case near Zürich (Lienert et al., 2015).

Figure 1: Simplified objectives hierarchy for sustainable wastewater infrastructure planning (adapted from Lienert et al., 2015). Grey shading: five main objectives with two sub-objectives each; lowest-level boxes are attributes with ranges from worst to best possible case in square brackets; abbreviations of main objectives (top row) in parentheses. WW = wastewater; No. = number; yr = year; p = person.

SWING weights were elicited with the standard procedure (e.g. Eisenführ et al., 2010). Our visualization with emoticons and questionnaires are given in the Supplementary Material (Section SI-1.1.3; Box SI-1, SI-2, SI-5), and one example in Appendix A. For the SMART/SWING variant, each respondent defines a most familiar reference objective (Box SI-6; for two sub-objectives no reference was needed), to which all other objectives are compared pairwise (Mustajoki et al., 2005). Respondents then improve the more important (sub-)objective from the worst to the best state, given the ranges of the case study (or choose both as equally important; Box SI-3, SI-7). They then define the strength of importance with a slider (Box SI-4, SI-8). In our variant, respondents used a slider and assigned 10 points to the least important objective, the reference or the other; in SMART/SWING, the reference received 10 points. We also gave verbal cues

(Huizingh & Vrolijk, 1997) from 10 to 90 (Box SI-4, SI-8). The scores were normalized (Appendix B).

(C) Feedback questions designed to assess the perceived difficulty of each method: four of five questions were identical for both methods: e.g. “How difficult was it overall to answer the questions?”, “How strongly did you consider the worst possible and best possible state of the two objectives (...)?”, and reasons for the difficulties (Tab. SI-9). Some questions were method-specific; e.g. “How difficult was it to”: “(...) rank alternatives with objectives you want to improve (SWING)?”

(D) Explanatory variables as possible additional explanations for different preferences among respondents and for preference stability: the questions concerned demographic variables (gender, having children, nationality, age, education), environmentally friendly behavior (public transport, financial contribution to environmental organizations, the respondent’s relation to water (living nearby, frequency of visits), and mood (Tab. SI-4). The second survey questionnaires were much shorter: we did not repeat the knowledge and feedback questions, but asked about new knowledge or experience between surveys (Tab. SI-10; “new experience”). The order of sub-objectives was randomized.

3.3 Statistical tests

Although we randomly assigned the respondents to one of the two methods, a systematic pattern may emerge purely by chance, such as significantly more men being assigned to SWING. We therefore tested for differences in descriptive variables between different groups, e.g. respondents assigned to SWING or the SMART/SWING variant, from the public or Eawag, or those who completed only the first or both surveys. We compared categorical variables with Chi-square tests and ordinal/continuous variables with t-tests (or used the Mann Whitney and Wilcoxon signed rank test if our assumptions were not met; Field, 2009). We used ANOVA for >2 categories and regression for continuous variables (age, knowledge, experience). Additionally, we tested the effect of explanatory variables with hierarchical regression. Hereby, to control for method, knowledge, experience, and new experience were entered first, while the method was entered in the second block. Statistical analyses were conducted with SPSS 21.0. The MCDA was implemented in R (www.R-project.org) using the R package utility (Reichert et al., 2013).

3.4 Weight profiles and preference stability of two elicitation methods

Using weights from the first survey, we searched for patterns in the importance of the (sub-)objectives within given range intervals. We then analyzed if there was a difference between SWING and the SMART/SWING variant, if preferences changed over one month, and if so, why. Specifically, we explored:

1. Differences in weights elicited with SWING and the SMART/SWING variant;
2. Differences in stability of weights over time, i.e. between the first and second surveys;
3. The effect of explanatory variables on weight stability in a regression model: method, knowledge, experience, age, education, and new experience gained between the first and second surveys;
4. The influence of strong preferences on weight stability, i.e. difference between respondents who gave extreme weights and those with a more even weight distribution.
5. The perceived difficulty of the SWING questionnaire and the SMART/SWING variant.

We tested for differences between the two methods (point 1) and the two surveys over time (point 2) with t-tests, and analyzed the effect of explanatory variables with hierarchical regression (point 3; see Section 3.3). We needed indicators to test the weight stability (point 2) and used the following four: (i) proportion of rank reversals of sub- and (ii) main objectives; (iii) sum of absolute differences (SAD) in weights of sub- and (iv) main objectives. (i) A rank reversal occurred if the first ranked sub-objective (of two/ main objective) in the first survey ranked second or equal in the second survey, or if an indifference statement was changed to a preference. The proportion of rank reversals for (i) and (ii) ranged from 0 (no rank reversals) to 1 (Appendix C). For the SAD (iii, iv), we summed up all changes in the normalized weights between the first and second surveys (Bottomley et al., 2000). SAD ranged from 0 (no change in weights of (iii) sub-objectives) to 10; and 0 (no change for (iv) main objectives) to 5, respectively.

We addressed strong preferences (point 4) on the basis of a line of reasoning (Supplementary Material; Section SI-2.6.4). We tested for correlations between the linear coefficients of the weight vectors in the first survey and the stability indicators

(rank reversals, SAD). We then clustered the respondents into “extreme weights” vs. more “even weight distribution” (using thresholds⁵) and tested the preference stability for SWING or the SMART/SWING variant. The questions used to assess the perceived difficulty (point 5) are given in Section 3.2.

3.5 Do different methods give the same ranking of alternatives?

We analyzed whether differences in weights affect the resulting MCDA values and ranks of alternatives. For this, we used the MCDA model and data from a real wastewater infrastructure case near Zürich that was part of the SWIP project. The case study was analyzed in much more detail than is possible in this paper, and included a combination of MCDA with scenario analysis as well as careful uncertainty modeling and sensitivity analyses. The application to the wastewater infrastructure system and more details concerning the MCDA method applied here are given in Zheng et al. (subm.). A separate application to the water supply system is published in Scholten et al. (2015). Here, we chose a subset of six wastewater alternatives that differ in their wastewater, drainage and stormwater systems, rehabilitation strategies and organizational structures (Appendix D; same alternative labels as Zheng et al., subm.). These alternatives cover the extremes and provided the most discriminating results in the earlier MCDA. They include variants of the current system with large central wastewater treatment plants: the status quo with improved rain infiltration implemented over time (A8a; Appendix D); a technically advanced variant to remove micropollutants with good pipe maintenance (A2); a central cheaper system with poor maintenance (A9). The fully decentralized systems cover an inexpensive low-tech system (A5), a high-tech system with good environmental performance (A7) and a semi-decentralized system between the extremes (A8b).

For each of the selected six alternatives, we predicted how well they achieve the ten sub-objectives (the average attribute levels are given in Tab. SI-3; we also give the best and worst possible states). We used the additive value model with linear value functions for every attribute (Keeney & Raiffa, 1976). The MCDA was performed for the group of

⁵ For example, we compared respondents who had a weight of 0.4 or higher for one main objective in the first or second survey vs. those with weights <0.4 for all main objectives in both surveys. We chose 0.4, because very few people had weights >0.4 for SWING; for the SMART/SWING variant, very few had <0.4.

respondents using SWING in the first (N=158) or second survey (N=94); or the SMART/SWING variant (N=156 first; N=106 second survey) respectively. For each group, we implemented the average weight over all individuals. We additionally tested whether the best alternative was consistent between the first and second survey for each individual respondent. For this purpose, we introduced their personal weights into the MCDA and used Kendall's tau (1938) correlation coefficient to analyze the ranking of the alternatives.

4 Results

4.1 Sample description

We sent 708 invitations to the public and 807 to Eawag (for details, see Section SI-2.1). Our sample (Tab. 1) corresponds well with the average Swiss population (see Section SI-2.2). To ensure unbiased sampling, we compared the demographic and descriptive variables of the groups; we found very few statistically significant differences (see Section SI-2.4).

Table 1: Number of respondents and completion rates

	Number of responses		
	Survey 1	Survey 2	Completion rate
Public			
SWING	123	80	64%
SMART/SWING variant	126	93	74%
Eawag			
SWING	35	14	40%
SMART/SWING variant	30	13	43%
TOTAL	314	200	63%

4.2 Weight profiles

In the overall sample, the single objectives 'protection of water and other resources' and 'safe wastewater disposal' received higher weights than 'intergenerational equity', 'high

social acceptance', and 'low costs', alone (Tab. 2; also see Discussion Fig. 3), on the basis of the case study ranges. The sub-objectives followed a similar pattern.

Table 2: Average weights of main and sub-objectives of all respondents who completed the first survey (N=314). We show standard errors (SE) and deviations (SD). WW= wastewater.

Main objectives	Average	SE	SD
Intergenerational equity	0.17	0.01	0.13
Protection of water and other resources	0.27	0.01	0.16
Safe wastewater disposal	0.26	0.01	0.15
High social acceptance	0.13	0.01	0.10
Low costs	0.16	0.01	0.13
Sub-objectives	Average	SE	SD
Low future rehabilitation burden until 2050	0.09	0.00	0.08
Flexible system adaptation	0.08	0.05	0.08
Good chemical state of watercourses	0.16	0.01	0.12
Recovery of phosphate	0.11	0.01	0.11
Number of gastrointestinal infections through WW	0.13	0.01	0.10
Sufficient drainage capacity of drainage system	0.13	0.01	0.11
Low disturbance by unnecessary construction	0.08	0.00	0.07
Low time demand for end user	0.06	0.00	0.07
Low annual costs	0.08	0.00	0.08
Low cost increase	0.08	0.00	0.08

4.2.1 Differences between public and Eawag respondents

The public (N=249) and Eawag respondents (N=65) had similar weights, with two exceptions in the first survey. Eawag gave statistically significantly higher weights to 'protection of water and other resources' (0.33, $t = -4.2$, $p < .001$) and lower ones to 'low costs' (0.11, $t = 4.4$, $p < .001$) than the public (0.25 and 0.17 respectively; Fig. 2; Tab. SI-17). The corresponding sub-objectives 'good chemical state' and 'no infections' received higher weights from Eawag, whereas 'drainage capacity', 'annual costs', and 'cost increase' were weighted more highly by the public. Follow-up tests indicated that the

higher weights for Eawag were not related to their higher education level (Section SI-2.5.1).

Figure 2: Average weights given to main objectives by the public (white bar) and Eawag respondents (shaded bar). Error bars= 95% confidence interval

4.2.2 Differences in relation to demographic and other explanatory variables

Gender and age were not statistically significantly correlated with the weights (Tabs. SI-21; SI-23). Parents gave slightly higher weights to ‘intergenerational equity’ than those without children (Tab. SI-21; Fig. SI-1). University graduates gave statistically significantly higher weights to ‘protecting water and other resources’ than people from vocational schools (Tabs. SI-24–25; Fig. SI-2). However, university graduates gave lower weights to ‘costs’ (Tab. SI-26). The frequency of visiting or living near water, using public transport for ecological reasons and knowledge were not related to the weights. People who often give financial support to environmental organizations gave statistically significantly higher weights to ‘protecting water and other resources’ (Tabs. SI-27–28; Fig. SI-3).

4.3 Preference stability of two elicitation methods

4.3.1 Differences in weights elicited with two methods and stability over time

‘Protection of water and other resources’ had higher weights if respondents used the SMART/SWING variant (weight 0.30, SWING: 0.24; $t = -3.1$, $p = .002$; Tab. SI-34; Fig. SI-5); SWING respondents gave higher weights to ‘social acceptance’ (0.15 vs. 0.12; $t = 2.54$, $p = .012$). Users of the SMART/SWING variant had higher rank reversals and SAD for the sub- and main objectives (Tab. 3).

Table 3. Comparison of the proportion of rank reversals (RR) and sum of absolute differences (SAD) in the weighting of main and sub-objectives in a retest survey after one month (t-test). RR ranged from 0 (no RR) to 1, SAD from 0 to 5 (main objectives), or from 0 to 10 (sub-objectives).

Indicators of stability	SWING (N=94)		SMART/SWING variant (N=106)		t	Significance
	Average	Standard error	Average	Standard error		
RR sub-objectives	0.43	0.03	0.51	0.02	-2.21	.028
RR main objectives	0.32	0.02	0.50	0.02	-6.04	<.001
SAD sub-objectives	1.45	0.08	2.68	0.13	-7.73	<.001
SAD main objectives	0.38	0.03	0.77	0.04	-8.30	<.001

4.3.2 Explanatory variables and observed weight profiles over time

(i) Difference between two elicitation methods

The hierarchical regression analysis showed a statistically significant to highly significant difference between the two elicitation methods for all four measures of preference stability (Tab. 3). In three of four cases, the correlation between method and the rank reversals or SAD's of the weights was considerably greater than for the other explanatory variables (Beta= .39, .48, .49; exception: rank reversal of sub-objectives, Beta= .14; Tab. 4).

(ii) Knowledge and experience

Knowledge was statistically significant to highly significant for three of four predictors of preference stability (not for SAD main objectives; Tab. 4; see similar results for only the public, who had less knowledge than Eawag; Tab. SI-35). As knowledge increased, so did preference stability. Experience was not correlated with preference stability.

(iii) Age and education

Age was statistically significantly positively correlated with three of four measures of preference stability (not for RR of sub-objectives): older respondents had less stable preferences (Tabs. 4; SI-35). There was a statistically significant interaction between method and age (Fig. SI-6) for the SAD of sub-objectives: SAD increased with

increasing age for the SMART/SWING variant, but remained almost stable for SWING. There was no relationship between preference stability and education.

Table 4: Results of four hierarchical regression analyses for four indicators of preference stability in a retest survey after one month: rank reversals and sum of absolute differences in weights of main and sub-objectives (Appendix C). Explanatory variables: elicitation method, wastewater knowledge/ experience, age, education level, and new experience between first/ second survey. Graduates from primary/ secondary school were excluded (very few data; Tab. SI-7). Beta is a standardized coefficient (normalized to [0, 1]); *P*-values in parentheses; ns = not significant.

	Rank reversals		Sum of absolute differences	
	Sub-objectives	Main objectives	Sub-objectives	Main objectives
Method	.144 (.034)	.386 (<.001)	.482 (<.001)	.492 (<.001)
Knowledge	-.269 (<.001)	-.138 (.038)	-.132 (.035)	-.046 (ns)
Experience	-.024 (ns)	-.073 (ns)	.072 (ns)	.030 (ns)
Age	.069 (ns)	.187 (.005)	.132 (.034)	.186 (.003)
Education	-.127 (.071)	-.024 (ns)	-.097 (ns)	-.090 (ns)
New experience	.150 (.052)	.158 (.032)	.129 (.062)	.138 (.046)
R ²	.154	.229	.324	.325

(iv) New experience

The hierarchical regression analysis gives no clear indication as to whether new experience between the first and second survey is an important explanatory variable: there was a statistically significant effect for both preference stability measures for the main objectives (Tab. 4); respondents with more new experience had less stable preferences. However, the Beta value was relatively low, and the effect was only marginally significant for the sub-objectives. There was a marginally significant interaction between method and new experience: as new experience increased, SAD of sub-objectives also increased for the SMART/SWING variant, but not for SWING (Fig. SI-7).

(v) Influence of strong preferences

In the regression analysis of the main objectives, there was no difference in rank reversals for respondents with more extreme or balanced weights. For SAD, more

extreme weights were correlated with less stable preferences (i.e. statistically significant higher SAD; Tab. 5).

The cluster analysis produced the same result: there was no difference in the proportion of rank reversals for respondents with more extreme preferences compared to those with a more even weight distribution. However, those with more extreme weights had statistically significantly less stable SAD compared with those with a more even weight distribution. For reasons of space, we present the details in the Supplementary Material (Tabs. SI-37–41).

Table 5: Regression analysis to test effects of extreme preferences with linear coefficients of weight vectors (Coeff.) for two weight elicitation methods. Indicators of preference stability: proportion rank reversals and sum of absolute differences (SAD) in weights of main objectives.

	SWING (N=94)			SMART/SWING variant (N=106)		
	Coeff.	SE	P	Coeff.	SE	P
Rank reversals	-0.218	0.852	.798	-0.573	0.590	.333
SAD	2.917	0.839	.001	4.650	1.151	<.001

4.3.3 Perceived difficulty of SWING questionnaire and SMART/SWING variant

A quarter of all respondents ticked that the questionnaire was “somewhat easy”, “easy”, or “very easy”, almost 70% of whom used the SMART/SWING variant (Tabs. 6; SI-9–10; Fig. SI-4). Although 34% of those using the SMART/SWING variant perceived it as easy (sum of above classes), the figure was only 15% for SWING. Consequently, SWING differed highly significantly from the SMART/SWING variant in statistical terms: SWING had an average rating of 4.1, but the SMART/SWING variant only 3.7 (Tab. SI-31; same both surveys: Tab. SI-32), on a scale of 1 to 6, where 3= “somewhat easy”, and 4= “somewhat difficult” (Tab. 6). We point out that the differences between 4.1 for SWING and 3.7 for the SMART/SWING variant may seem small, but this is a result of averaging the six classes. Differences within classes were substantial, for instance, only 5% of people using SWING ticked “somewhat easy”, but the figure was 14% for those using the SMART/SWING variant (Tab. 6). SWING users felt statistically significantly less certain about their answers, but again with small average differences between groups (e.g. 3.3 vs. 3.5 on a scale of 1 to 5; where 3= “fairly unsure” and 4= “highly sure”; Tabs.

6; SI-9; SI-31). There were no differences between methods for considering ranges of objectives (Tabs. SI-31–32).

Table 6: Feedback on the difficulty of the questionnaires of two weight elicitation methods and certainty about answers from respondents who completed the first survey (N=314). Number (N) and % of answers in each category (for more feedback questions, see Tab. SI-9).

Questions	Answer levels	SWING (N=158)		SMART/SWI NG variant (N=156)	
<i>How difficult was it overall to answer the questions?</i>					
	1 = Very easy	2	1%	8	5%
	2 = Easy	14	9%	24	15%
	3 = Somewhat easy	8	5%	22	14%
	4 = Somewhat difficult	86	54%	68	44%
	5 = Difficult	33	21%	26	17%
	6 = Very difficult	15	10%	8	5%
<i>How certain do you feel about your answers to the previous questions?</i>					
	1 = Extremely unsure. I am 99% unsure	4	3%	3	2%
	2 = Highly unsure. I am 75% unsure	27	17%	12	8%
	3 = Fairly unsure. I am 50% unsure	54	34%	58	37%
	4 = Highly sure. I am 25% unsure	65	41%	71	46%
	5 = Extremely sure. I am 1% unsure	8	5%	12	8%

4.4 Do different methods give the same ranking of alternatives?

As shown above, we found clear differences between weights of SWING and the SMART/SWING variant (Section 4.3). We then analyzed whether the different weights affect the total value and ranking of the six wastewater alternatives. For this purpose, we first introduced the average weights of groups into an MCDA model. Despite significant differences in weights, there were no statistically significant differences between the resulting MCDA values of the alternatives for the following groups: first survey SWING vs. SMART/SWING variant; first vs. second survey with SWING; and first vs. second

survey with SMART/SWING variant (Tab. 7). One exception was a minor difference for alternative A5 in the second survey of SWING vs. SMART/SWING variant.

Consequently, the ranking of alternatives was identical for all four groups using average group weights. The best-ranked alternative was A7 in all cases. We compared these result with weights obtained from ten stakeholders in face-to-face interviews (Zheng et al., subm.; calculation of predictions of alternatives, see there). To calculate the average values reported in Table 7 (last column), we introduced the weights of the ten stakeholders into the same MCDA model as the online surveys (i.e. linear marginal value functions, additive aggregation, no uncertainty). Again, alternative A7 ranked best.

Table 7: MCDA results. Total values of six decision alternatives (A2–A9; Appendix D) after implementing average weights of groups using two weight elicitation methods in two survey rounds, one month apart. Last column: overall average value with weights from 10 stakeholder interviews (Zheng et al., subm.). Values range between [0,1]; 0= objective ‘sustainable wastewater infrastructure’ is not at all achieved; 1= objective is fully achieved. N= sample size. Bold = highest values, above 0.7.

Alternative	SWING		SMART/SWING variant		Inter- view
	Round 1	Round 2	Round 1	Round 2	
	(N=158)	(N=94)	(N=156)	(N=106)	
A2 Central high tech	0.667	0.656	0.664	0.673	0.741
A5 Decentral low tech	0.581	0.572	0.595	0.606	0.639
A7 Decentral high-tech nutrient recovery	0.805	0.804	0.806	0.807	0.807
A8a Status quo (central), stormwater retention	0.638	0.636	0.644	0.643	0.800
A8b Decentral high tech	0.621	0.621	0.630	0.627	0.798
A9 Central privatized, minimal maintenance	0.535	0.526	0.558	0.561	0.623

We then analyzed for each individual whether changes in weights between the first and second survey would lead to different rankings of alternatives. For this purpose, we

correlated the rankings of the alternatives in the first and second surveys. The average Kendall's tau correlation coefficient of rankings between the two surveys was 0.625 (SD=0.325). We found a difference between the two weight elicitation methods. Using SWING, the best ranked alternative was always A7 in the first and the second surveys after one month, with one exception, i.e. 93 of 94 people had A7 on the first rank (=99%; see Discussion Fig. 4; Tab. SI-42). The rankings of alternatives were less stable for respondents using the SMART/SWING variant: only for 78% (83 of 106) of them did the *same* alternative rank best in the second survey, as in the first. This was reflected in the average correlations between the two surveys using the weights for individuals, where SWING yielded statistically significantly more consistent rankings (Kendall's tau: 0.730) than the SMART/SWING variant (Kendall's tau: 0.532; $N=200$, $t=4.612$, $p<.001$; Fig. SI-10). But alternative A7 also performed best for the majority of individuals using the SMART/SWING variant: in the first survey, A7 was best in 87% of cases, and for 88% of the respondents in the second survey. Second best was alternative A2, namely for 10% of respondents in the first survey, and 8% in the second (Fig. 4; Tab. SI-42).

5 Discussion

5.1 Weight profiles

5.1.1 Similar weights for different stakeholder groups

The use of different elicitation procedures and inclusion of different stakeholder groups can increase our confidence in the resulting weight profiles. We carried out an online survey among the Swiss public and water experts using two different weight elicitation methods and repeated the survey with both methods after one month. The weight distributions of the five main objectives (Tab. 2) important for good wastewater infrastructures were very similar to earlier face to face interviews with 20 Swiss stakeholders (Scholten et al., 2015; Zheng et al., *subm.*): in all studies, 'protection of water and other resources' and 'safe wastewater disposal' (resp. 'good water supply') received higher weights than 'intergenerational equity', 'high social acceptance', and 'low costs' (Fig. 3). The same attribute ranges were used in all studies. Current wastewater costs in the study region near Zürich are CHF 289 per person and year (Tab. SI-3). Allowing pipes to deteriorate (alternative A9; Appendix D), for example, would

reduce costs to CHF 157 p⁻¹yr⁻¹, but on average people did not find this substantial enough to trade off against the negative effects. The importance of water protection is in line with Swiss Willingness to Pay (WTP) studies (Logar et al., 2014; Veronesi et al., 2014).

Figure 3: Average weights (and standard errors) of the five main objectives elicited in multiple independent studies. From the left: interviews with ten water supply stakeholders (SH; dark grey; Scholten et al., 2015); interviews with ten wastewater SH (light grey; Zheng et al., subm.); first online survey, SWING (diagonal lines); SMART/SWING variant (dots).

We found very few statistically significant relationships between the weights for the objectives and demographic or other explanatory variables (Section 4.2.2). We generally conclude that the average weights of the five main objectives are similar, both when they were elicited in personal interviews and online. However, we found a difference in weights with respect to the elicitation method (Sections 5.1.2 and 5.2) and therefore also verified whether this would change the outcome of a real decision (Section 5.3).

5.1.2 SMART/SWING variant caused a larger spread of weights

AHP is criticized due to inconsistencies (Belton, 1986; Salo & Hämmäläinen, 1997) and a larger spread of weights (Pöyhönen & Hämmäläinen, 2001; and references therein). Our result of a larger spread of weights for respondents using the SMART/SWING variant

agrees with this, because we used an AHP-type verbal scale (Huizingh & Vrolijk, 1997) only in the SMART/SWING variant questionnaire, but not in SWING. Firstly, we found that respondents using the SMART/SWING variant gave statistically significantly higher weights of 0.30 to the most important objective of 'protection of water and other resources', while SWING respondents gave only 0.24 (Section 4.3.1). At the same time, users of the SMART/SWING variant gave statistically significantly lower weights to the least important objective of 'high social acceptance' (0.12), while SWING respondents gave this a slightly higher weight (0.15). Secondly, we tested whether the SMART/SWING variant leads to a higher spread of weights than SWING by sorting each respondent's weights from highest to lowest and regressing the weights on the ranking to capture the linear component (as in Doyle et al., 1997; Bottomley et al., 2000; details see Section SI-2.6.4). Again, we found some statistically significant support that the SMART/SWING variant leads to a larger spread of weights. However, these results are not quite evident because weight differences between the two methods seem small. Moreover, one would expect an effect for the second important objective, which was not the case. It thus remains unclear whether the SMART/SWING variant would also lead to a larger spread of weights in other applications.

5.2 Preference stability of two elicitation methods over time

5.2.1 SWING produced more stable weights

The literature provides ample evidence that different elicitation methods produce different weights or risk attitudes (see review Section 2). We found that the method can also affect preference stability over time, which follows the scarce literature (Bottomley & Doyle, 2001, 2013; Bottomley et al., 2000): respondents using the SMART/SWING variant had statistically significantly more changes in weights in the second survey for all four preference stability indicators (Section 4.3.1). The average rank reversal of weights was 51% (sub-objectives) and 50% (main objectives) for the SMART/SWING variant, but only 43% (sub-objectives) and 32% (main objectives) with SWING (Tab. 3). Thus there was a substantial difference between the two methods for the proportion of rank reversals. The weight difference, which we measured as Sum of Absolute Differences (SAD) were smaller, e.g. for main objectives 0.38 (SWING) vs. 0.77 (SMART/SWING variant) on a scale of 0 (no difference first/ second survey) to 5. In combination, our

results for the SAD and rank reversals suggest a systematic difference between methods that cannot be explained by any other variable such as demography, since the two groups did not differ in this respect. The questionnaire was also very similar for both methods apart from the fact that an AHP rating scale was used in the SMART/SWING variant. We discussed above that the AHP rating scale might cause a larger spread of weights: this can affect the *weighting*, which we measured as SAD. The reason is that the important objectives receive a higher weight if there is a larger spread of weights. Therefore, any differences in weights between the first and second survey will also be larger, i.e. the SAD will be larger for the SMART/SWING variant. This in turn leads to the conclusion that the SMART/SWING variant has a lower preference stability compared to SWING if we rely only on the SAD as an indicator. However, the AHP scale cannot account for more rank reversals. Even if there is a larger spread of weights for the SMART/SWING variant, the ranking of objectives remains the same. There should be no difference for the indicator rank reversals when comparing SWING with the SMART/SWING variant if the AHP rating scale and associated larger spread of weights is the only difference between the methods. We must therefore look for additional reasons that could explain the difference in preference stability between SWING and the SMART/SWING variant.

5.2.2 Effects of other explanatory variables on preference stability

According to the literature, it is difficult to foresee which variables best explain the stability of preferences. Some variables were only rarely correlated with preference stability, including socio-demographic variables and past experience, knowledge and expertise (see literature review in Section 2). Although learning processes are known to influence preferences, there is so far no evidence in the scant literature that new experience or the occurrence of external events affect preference stability. However, learning and new experience are important for environmental decision-making because they can be influenced by information from political actors or environmental agencies. Learning about environmental problems may help respondents to form more stable preferences that endure beyond a single situation, as requested by Gregory et al. (2012). In our surveys, none of the above explanatory factors were most important to explain preference stability: we used hierarchical regression to test the influence of explanatory

variables on the observed weight profiles over time. Quite clearly, the weight elicitation method was the most important predictor of preference stability (Section 4.3.2; Tab. 4). Additionally, preferences were more stable if respondents had more wastewater knowledge for three of four preference stability indicators and seemed to be less stable for older respondents. Experience and education were completely unrelated to preference stability. Our results are inconclusive for new experience gained between surveys. Although it makes intuitive sense that people update their judgment if they receive new information, we found a statistically significant effect only for the main objectives.

Moreover, there could be a difference in preference stability over time if people have strong preferences compared with those with a more even weight distribution (Hoeffler & Ariely, 1999; San Miguel et al., 2002). Our current results are inconclusive when we tested respondents with more extreme weight profiles versus those with a more balanced weight distribution. We found no difference in the preference stability for our indicator rank reversals. However, for the indicator SAD, respondents with stronger preferences had statistically significantly less stable weights compared to those with a more equal weight distribution (Section 4.3.2; Tab. 5; contrasting San Miguel et al., 2002). Currently, it is unclear whether this is a real effect or is again caused by an AHP spread of weights effect, as discussed above (Section 5.1.2): when weights are extreme, small relative deviations produce larger absolute differences, thus larger SAD.

In summary, the method per se was clearly the most important explanatory variable for preference stability over time. Additionally, we have strong indications that having more knowledge and expertise about wastewater-related issues is correlated with higher preference stability. The influence of other explanatory variables remains inconclusive, namely the age of respondents, new experience gained between surveys and the effect of strong preferences. But how can we explain the observed method effect?

5.2.3 SWING was perceived as slightly more difficult but produced more stable weights than the SMART/SWING variant

Hoeffler & Ariely (1999) postulate that preferences could be more stable if choices are hard than if they are easy. Additionally, they found that people who faced an easy choice were surer about their preferences than those facing a hard choice. The effect of

the difficulty of elicitation processes on preference stability requires further research, since we found only two other studies that mention this important topic (Brouwer, 2006; San Miguel et al., 2002).

Our results are in line with the scant literature. Both questionnaires were perceived as difficult (see Section 4.3.3). However, SWING was perceived to be more difficult (85% found it difficult) than the SMART/SWING variant (66% found it difficult; Tab. 6). This statistically significant difference can be explained by noting that the SMART/SWING variant was designed to ease elicitation: one suggestion to make elicitation easier is to allow respondents to choose any sub-objective as a reference (Mustajoki et al., 2005). Additionally, we gave verbal cues in our SMART/SWING variant to express the strength of preference. In contrast, SWING respondents had to rank hypothetical, extreme alternatives, which can be difficult (Box SI-1, SI-5). Moreover, these respondents felt a bit less certain about their answers: 54% using the SMART/SWING variant felt highly or extremely sure about the answers, but only 46% of subjects using SWING felt that (Tab. 6). In summary, we found the interesting effect that although the SWING questionnaire was perceived as more difficult, it produced more stable preferences over time compared with the easier SMART/SWING variant. We suggest that these findings be tested in other applications and with other elicitation methods.

5.3 Different methods only slightly changed the ranking of alternatives

Our results showed that preferences (weights) can change, for instance by using different elicitation methods, and over time. Despite this, our MCDA application to a real Swiss wastewater case consistently produced the same first-ranked alternative: alternative A7 received the highest values (Tab. 7) and best ranks in the first and second survey, irrespective of whether SWING or the SMART/SWING variant was used. Likewise, alternative A7 performed well in the ten earlier stakeholder interviews (Zheng et al., subm.; Fig. 4). But the alternatives A2, A8a, and A8b also performed well in the interviews (Tab. 7). We discuss the implications for wastewater management below.

Figure 4: Performance of six wastewater alternatives across independent weight-elicitation methods. Percentage of respondents for which alternative (A2–A9; Appendix D) ranked first, second, or third for: face-to-face interviews (Zheng et al., subm.; left bars), online elicitation with SWING (middle), or SMART/SWING variant (right bars; only first survey; second survey was not statistically significantly different).⁶

In important decisions that affect a long time period, the ranking of alternatives would ideally stay the same if elicitation is repeated. However, the literature showed that correlations between the first and second survey were often only ‘moderate’ (e.g. Jorgensen et al., 2004; Liebe et al., 2012; see Section 2.8). Our own results are fully in line with the correlation coefficients reported in the literature (average Kendall’s tau=0.63, SD=0.33). This relatively low correlation may lead to the premature conclusion that these preferences do not remain stable after one month, which might in turn challenge the decision itself. But if a decision-maker really has to select one of several possible

⁶ Reading example: The central high-tech alternative A2 (far left) ranked first (dark grey bars) for 0%, 0.6%, and 10% of the respondents in the interviews, SWING and SMART/SWING variant respectively. It ranked second (light grey bars) for 30%, 68%, and 46% respectively (same order), and third (diagonal bars) for 10% in all three methods (for details, see Tab. SI-42).

alternatives, the first-ranked alternative is more important than the ranking of all alternatives. In our case, the recommendation about the best alternative would change for about 12% of individuals due to instable weights over one month *if* the weights were elicited with the SMART/SWING variant (Kendall's $\tau=0.53$). For SWING, the best alternative remained the same for all 94 respondents (Kendall's $\tau=0.73$; one exception). This method effect cancelled out if we used average weights of groups: Rankings of alternatives were highly stable regardless of the elicitation method and time gap. Alternative A7 always had the highest values (Tab. 7). Nevertheless, we caution against using values that were averaged over a population sample in important environmental and policy decisions. We recommend that the results of individuals be carefully analyzed before a decision affecting the entire population is made (as do Schaafsma et al., 2014, who recommend using confidence intervals and sensitivity analyses).

The choice of the best-ranked alternative A7 would strongly affect the way we manage our wastewater. A7 is highly unconventional and assumes fully decentralized wastewater systems in houses by 2050. Urine separating toilets allow phosphate to be recovered as fertilizer for agriculture (Larsen et al., 2009). The good performance on this sub-objective 'phosph' seems to be the most important parameter for its high value of 0.81 (Tab. 7), since alternative A8b is similar, but does not recycle phosphorus (Appendix D). Alternative A7 performed better in the survey than in the interviews (Fig. 4). This is probably due to the use of a simplified hierarchy in the survey, with only ten sub-objectives instead of 19, whereby 'phosph' was weighted much more highly. In the interviews, the current central system (A8a) performed nearly as well as alternative A7. Can we now recommend alternative A7? It is reassuring that multiple elicitation procedures at different steps in time produced similar results, and A7 on average ranked first or second in 92% (SD=12%; Fig. 4; Tab. SI-42) of cases for all five procedures, which were online surveys or interviews. But the central alternatives A8a (rank 1, 2, or 3 for 73%, SD=19%) or A2 (65%, SD=15%) should also be discussed with decision-makers. Decentralized wastewater systems are proposed for sustainability reasons (e.g. Larsen et al., 2009) and are explored in some large engineering projects (e.g. in the

USA <http://www.renuwit.org/>; in Switzerland at Eawag⁷). This implies breaking the paradigm of the superiority of the central system, which is deeply rooted in engineering practice. Implementing a transition from central to decentralized systems requires strong evidence of their viability. We propose using complimentary methods, since a decision-maker would probably feel much more confident in choosing a highly unconventional decentralized wastewater system if MCDA interviews are backed up by public feedback.

6 Conclusions

To justify environmental and policy decisions, it is important to design elicitation procedures that produce stable results which reflect the stakeholders' preferences. Various factors influence preference statements, including the selected elicitation method. In our case, SWING produced statistically significantly different weights than the SMART/SWING variant. Moreover, in our study, the weights elicited with the SMART/SWING variant changed statistically significantly more often in the follow-up survey after one month compared to SWING. Despite its rigorous design, not all the limitations of such a study can be avoided (see Section SI-3.2). There is also ample scope for future research. To better understand preference formation and stability, we suggest using procedural and experimental analyses that address the following research questions:

1. Do popular elicitation methods such as SWING, SMART/SWING or trade-off (see e.g. Eisenführ et al., 2010) differ systematically in their elicited weight profiles, spread of weights and preference stability over time when applied to a variety of decision situations? The validity of online preference elicitation has been questioned (Marttunen & Hämäläinen, 2008). So virtual preference elicitation should also be compared with face-to-face procedures.
2. What are the reasons and underlying psychological mechanisms for the systematic differences between methods? For example, which explanatory variables influence preference formation and why? We found that preferences were more stable if respondents had more knowledge. To generalize this finding, it should be replicated in other settings. As another example, does the proposition

⁷ <http://www.eawag.ch/en/research/humanwelfare/wastewater/projekte/wings/>

of Hoeffler & Ariely (1999) hold that preferences are more stable if the decision task is more difficult? This question can be approached experimentally by constructing different types of questionnaires which allow easy choices or force people to think harder about a decision. Learning seems important, and observational studies could shed light on how learning processes evolve over time. Experiments allow us to study whether a systematic increase in information promotes more stable preferences. Is there a saturation point after which more information reduces preference stability, possibly because respondents resort to decision heuristics? Moreover, the influence of different information modes such as visual and verbal cues and interactive elicitation should be explored. As an example, does preference stability over time differ if the same information content is presented in different ways?

3. Environmental and policy decisions are especially sensitive to preference stability over a longer time range. Given that we can identify methods and processes that stabilize preferences, does this stabilizing effect wear off over time? Can this effect be avoided, for instance with targeted interventions at certain points in time?

Many of these research questions can be approached experimentally in decision labs and with the aid of students. However, to assure their generalization, laboratory results must be transferred to different types of real-world decision situations and stakeholders. Finally, for every concrete decision, it is highly relevant to identify whether possible differences in weight profiles and preference stability would result in a different ranking of alternatives in an MCDA. This need not necessarily be the case, as our study demonstrates: the first-ranked alternative was very stable, irrespective of preference differences relating to individual traits such as having more knowledge, the elicitation method (interviews, online elicitation with SWING or SMART/SWING variant), or a one-month time gap.

6 Acknowledgements

We thank Christoph Egger, Jonas Maria Eppler, Max Maurer, Peter Reichert, Lisa Scholten, Nele Schuwirth, and Anja Zorn for their valuable contributions, and the Swiss

Running title: Lienert, Duygan, & Zheng: Preference stability over time

National Science Foundation for funding via the National Research Program

"Sustainable Water Management" (NRP 61; www.nfp61.ch), project number:

406140_125901 / 1. We especially thank Eawag employees and public respondents for their participation. We are grateful for the patience and valuable comments of four anonymous reviewers and the editors Raimo Hämäläinen and Alberto Franco, all of whom greatly helped to improve the manuscript.

7 Supplementary Materials

Supplementary material for this article can be found in the online version at xxx

8 References

- Belton, V. (1986). A comparison of the analytic hierarchy process and a simple multi-attribute value function. *European Journal of Operational Research*, 26, 7-21.
- Belton, V., & Stewart, T. J. (2003). *Multiple Criteria Decision Analysis: An Integrated Approach*. Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Bliem, M., Getzner, M., & Rodiga-Lassnig, P. (2012). Temporal stability of individual preferences for river restoration in Austria using a choice experiment. *Journal of Environmental Management*, 103, 65-73.
- Borcherding, K., Eppel, T., & Von Winterfeldt, D. (1991). Comparison of weighting judgments in multi-attribute utility measurement. *Management Science*, 37, 1603-1619.
- Bottomley, P.A., Doyle, J.R. (2001). A comparison of three weight elicitation methods: good, better, and best. *Omega-International Journal of Management Science*, 29, 553-560.
- Bottomley, P.A., Doyle, J.R. (2013). Comparing the validity of numerical judgements elicited by direct rating and point allocation: Insights from objectively verifiable perceptual tasks. *European Journal of Operational Research*, 228, 148-157.
- Bottomley, P. A., Doyle, J. R., & Green, R. H. (2000). Testing the reliability of weight elicitation methods: Direct rating versus point allocation. *Journal of Marketing Research*, 37, 508-513.
- Brouwer, R. (2006). Do stated preference methods stand the test of time? A test of the stability of contingent values and models for health risks when facing an extreme event. *Ecological Economics*, 60, 399-406.
- Brouwer, R., Dekker, T., Rolfe, J., Windle, J. (2010). Choice certainty and consistency in repeated choice experiments. *Environmental & Resource Economics*, 46, 93-109.
- Cashman, A., & Ashley, R. (2008). Costing the long-term demand for water sector infrastructure. *Foresight*, 10, 9-26.
- Doyle, J. R., Green, R. H., & Bottomley, P. A. (1997). Judging relative importance: Direct rating and point allocation are not equivalent. *Organizational Behavior and Human Decision Processes*, 70, 65-72.
- Eisenführ, F., Weber, M., & Langer, T. (2010). *Rational Decision Making*: Springer.

- Field, A. (2009). *Discovering Statistics Using SPSS* (3rd Ed.): SAGE.
- Gregory, R., Lichtenstein, S., Slovic, P. (1993). Valuing environmental resources – a constructive approach. *Journal of Risk and Uncertainty*, 7, 177-197.
- Gregory, R., Failing, L., Harstone, M., Long, G., McDaniels, T., & Ohlson, D. (2012). *Structured decision making: A practical guide to environmental management choices*: Wiley-Blackwell Publishing.
- Hajkowicz, S. A. (2012). For the greater good? A test for strategic bias in group environmental decisions. *Group Decision and Negotiation*, 21(3), 331-344.
- Hajkowicz, S., & Collins, K. (2007). A review of multiple criteria analysis for water resource planning and management. *Water Resources Management*, 21, 1553-1566.
- Hämäläinen, R. P. (2015). Behavioural issues in environmental modelling – The missing perspective. *Environmental Modelling & Software*, 73, 244-253.
- Hämäläinen, R. P., & Alaja, S. (2008). The threat of weighting biases in environmental decision analysis. *Ecological Economics*, 68, 556-569.
- Hämäläinen, R. P., Luoma, J., & Saarinen, E. (2013). On the importance of behavioral operational research: The case of understanding and communicating about dynamic systems. *European Journal of Operational Research*, 228, 623-634.
- Hanley, G. P., Iwata, B. A., & Roscoe, E. M. (2006). Some determinants of changes in preference over time. *Journal of Applied Behavior Analysis*, 39, 189-202.
- Hey, J.D., Morone, A., Schmidt, U. (2009). Noise and bias in eliciting preferences. *Journal of Risk and Uncertainty* 39, 213-235.
- Hoeffler, S., & Ariely, D. (1999). Constructing stable preferences: A look into dimensions of experience and their impact on preference stability. *Journal of Consumer Psychology*, 8, 113-139.
- Huang, I. B., Keisler, J., & Linkov, I. (2011). Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. *Science of the Total Environment*, 409, 3578-3594.
- Huizingh, E., & Vrolijk, H. C. J. (1997). A comparison of verbal and numerical judgments in the analytic hierarchy process. *Organizational Behavior and Human Decision Processes*, 70, 237-247.
- Jacobi, S. K., & Hobbs, B. F. (2007). Quantifying and mitigating the splitting bias and other value tree-induced weighting biases. *Decision Analysis*, 4, 194-210.
- Janis, I. L. (1972). *Victims of Groupthink*. New York: Houghton Mifflin.
- Jorgensen, B.S., Syme, G.J., Smith, L.M., Bishop, B.J. (2004). Random error in willingness to pay measurement: A multiple indicators, latent variable approach to the reliability of contingent values. *Journal of Economic Psychology* 25, 41-59.
- Keeney, R. L., & Raiffa, H. (1976). *Decisions with Multiple Objectives: Preferences and Value Tradeoffs*. New York: Wiley.
- Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 1938, 81-93.
- Larsen, T. A., Alder, A. C., Eggen, R. I. L., Maurer, M., & Lienert, J. (2009). Source separation: Will we see a paradigm shift in wastewater handling? *Environmental Science & Technology*, 43(16), 6121-6125.
- Liebe, U., Meyerhoff, J., Hartje, V. (2012) Test-retest reliability of choice experiments in environmental valuation. *Environmental & Resource Economics*, 53, 389-407.
- Lienert, J., Schnetzer, F., & Ingold, K. (2013) Stakeholder analysis combined with social network analysis provides fine-grained insights into water infrastructure planning processes. *Journal of Environmental Management*, 125, 134-148.

- Lienert, J., Scholten, L., Egger, C., & Maurer, M. (2015). Structured decision-making for sustainable water infrastructure planning and four future scenarios. *EURO Journal on Decision Processes* 3(1-2), 107-140. (special issue on Environm. Decision Making).
- Logar, I., Brouwer, R., Maurer, M., & Ort, C. (2014). Cost-benefit analysis of the Swiss national policy on reducing micropollutants in treated wastewater. *Environmental Science & Technology*, 48, 12500-12508.
- Marttunen, M., & Hämäläinen, R. P. (2008). The decision analysis interview approach in the collaborative management of a large regulated water course. *Environmental Management*, 42, 1026-1042.
- Maurer, M., & Lienert, J. (2014). Sustainable water infrastructure planning for an uncertain future (in German, English, French). *Eawag news* 02/ April 2014, 1-8. http://www.eawag.ch/medien/publ/news/2014_na_17/index?clear_lang=1
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics*. New York: Academic Press.
- Mendoza, G. A., & Martins, H. (2006). Multi-criteria decision analysis in natural resource management: A critical review of methods and new modelling paradigms. *Forest Ecology and Management*, 230, 1-22.
- Milly, P. C. D., Betancourt, J., Falkenmark, M., Hirsch, R. M., Kundzewicz, Z. W., Lettenmaier, D. P., & Stouffer, R. J. (2008). Climate change – Stationarity is dead: Whither water management? *Science*, 319 (5863), 573-574.
- Montibeller, G. & von Winterfeldt, D. (2015). Cognitive and motivational biases in decision and risk analysis. *Risk Analysis* 35, 1230-1251.
- Morton, A., & Fasolo, B. (2009). Behavioural decision theory for multi-criteria decision analysis: a guided tour. *Journal of the Operational Research Society*, 60, 268-275.
- Mustajoki, J., Hämäläinen, R. P., & Salo, A. (2005). Decision support by interval SMART/SWING-incorporating imprecision in the SMART and SWING methods. *Decision Sciences*, 36, 317-339.
- OECD, (2012). http://stats.oecd.org/Index.aspx?DatasetCode=SNA_TABLE1. Data extracted on 28 August 2012, 12:51 UTC(GMT).
- Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research – A constructive processing perspective. *Annual Review of Psychology*, 43, 87-131.
- Pöyhönen, M., & Hämäläinen, R. P. (2001). On the convergence of multiattribute weighting methods. *European Journal of Operational Research*, 129, 569-585.
- Reichert, P., Schuwirth, N., & Langhans, S. (2013). Constructing, evaluating and visualizing value and utility functions for decision support. *Environmental Modelling & Software*, 46, 283-291.
- Salo, A. A., & Hämäläinen, R. P. (1997). On the measurement of preferences in the analytic hierarchy process. *Journal of Multi-Criteria Decision Analysis*, 6, 309-319.
- San Miguel, F. S., Ryan, M., & Scott, A. (2002). Are preferences stable? The case of health care. *Journal of Economic Behavior and Organization*, 48, 1-14.
- Schaafsma, M., Brouwer, R., Liekens, I., De Nocker, L. (2014). Temporal stability of preferences and willingness to pay for natural areas in choice experiments: A test-retest. *Resource and Energy Economics*, 38, 243-260.
- Scholten, L., Schuwirth, N., Reichert, P., & Lienert, J. (2015). Tackling uncertainty in multi-criteria decision analysis – An application to water supply infrastructure planning. *European Journal of Operational Research*, 242(1), 243-260.

- Schuwirth, N., Reichert, P., & Lienert, J. (2012). Methodological aspects of multi-criteria decision analysis for policy support: A case study on pharmaceutical removal from hospital wastewater. *European Journal of Operational Research*, 220, 472-483.
- Slovic, P. (1995). The construction of preference. *American Psychologist* 50, 364-371.
- Van Ittersum, K., Pennings, J. M. E., Wansink, B., & Van Trijp, H. C. M. (2007). The validity of attribute-importance. measurement: A review. *Journal of Business Research*, 60, 1177-1190.
- Veronesi, M., Chawla, F., Maurer, M., & Lienert, J. (2014). Climate change and the willingness to pay to reduce ecological and health risks from wastewater flooding in urban centers and the environment. *Ecological Economics*, 98, 1-10.
- Weber, M., & Borchering, K. (1993). Behavioral-influences on weight judgments in multiattribute decision-making. *European Journal of Operational Research*, 67, 1-12.
- Zeisberger, S., Vrecko, D., Langer, T. (2012). Measuring the time stability of Prospect Theory preferences. *Theory and Decision*, 72, 359-386.
- Zheng, J., Egger, C., & Lienert, J. (subm.). A scenario-based MCDA framework for wastewater infrastructure planning.

Appendix A: Visualization of weight elicitation

Elicitation of weights (scoring) to compare two sub-objectives with SWING as one example. Full details of both methods see Supplementary Information (Box SI-1–8). First, the ranges of the two attributes were described in words (best and worst case) along with the Status quo, and visualized with emoticons. People were then asked to rank the alternatives (Box SI-1). Based on rankings, people then scored the alternatives:

Scoring of alternatives

You ranked the **Alternative A** as, 1st. This alternative gets the highest score, **100**. The **Null Alternative** with all objectives at their worst level gets the lowest score of **0**.

Now we will ask you to score **Alternative B** from 0 to 99 according to its level of importance. Please use the slider to score Alternative B below. You don't need to score Alternative A and Null alternative.

Alternative	Score	Instruction	Low future rehabilitation burden until 2050	Flexible system adaptation
Alternative A	100	Please do not change	Green (Smiley)	Red (Frowny)
Alternative B	?	Please score it	Red (Frowny)	Green (Smiley)
Null Alternative	0	Please do not change	Red (Frowny)	Red (Frowny)

Slider: 0 25 50 75 100

Appendix B: Normalization of weights

For the SMART/SWING variant, the elicited scores from 10 to 90 were normalized: the ratios between the score and 10 was regarded as the ratio between the weights of the two objectives. The weights were hierarchically calculated from the ratios with an additional constraint that they sum up to unity. For instance, if the objective ‘low rehabilitation burden’ (‘rehab’) was judged “somewhat more important” than ‘flexible system adaptation’ (‘flex’), the score 30 (10, resp.) was assigned to ‘rehab’ (‘flex’, resp.). The weights of ‘rehab’ and ‘flex’ were calculated as 0.75 and 0.25 using two constraints: $w_{\text{rehab}} + w_{\text{flex}} = 1$ and $w_{\text{rehab}}/w_{\text{flex}} = 30/10$.

Appendix C: Indicators of preference stability

Proportion of rank reversals of sub-objectives

The highest possible number of rank reversals is five (five pairwise comparisons for ten sub-objectives; proportion=1). The lowest proportion is 0 (no rank reversals).

Proportion of rank reversals of main objectives

Respondents using SWING ranked all five objectives, but for the SMART/SWING variant, respondents chose a reference objective with which the other four objectives were compared. This did not produce a ranking. Therefore, for both methods, the weights given to main objectives were used to derive the rankings.

We computed the proportion of rank reversals with a matrix of five pairwise comparisons for each main objective (ten different comparisons; Tab. C-2). For each comparison, we determined whether objective x was ranked higher than objective y (>), the same (=), or lower (<). We calculated the proportion of changes. For the example in Table C-1, eight of ten comparisons were ranked differently in the second survey (Tab. C-2, right), which yields a proportion 0.8 rank reversals. The change in rankings can also be analyzed by correlation statistics, but we preferred this approach because it allowed comparing the rank reversals of sub-objectives with those of the main objectives.

Table C-1. Example of ranking of main objectives (A–E) in the first and second survey

Main objectives	First survey ranking	Second survey ranking
Intergenerational equity (A)	3	2
Protection of water and other resources (B)	1	5
Safe wastewater disposal (C)	2	3
High social acceptance (D)	4	1
Low costs (E)	4	4

Table C-2. Example of comparison of the ranking of the five main objectives A–E in the first (left) and second survey (right). Objective x can be ranked higher than objective y (>), equal (=), or lower (<). Changes between the first and second survey are highlighted with shading.

First survey						Second survey					
	A	B	C	D	E		A	B	C	D	E
A		<	<	>	>	A					>
B			>	>	>	B					
C				>	>	C					>
D					=	D					
E						E					

Sum of absolute differences (SAD) of main and sub-objectives

Table C-3. Calculation of sum of absolute differences (SAD) for main objectives. Example with different weights assigned to the main objectives in the first and second survey.

Main objective	First interview	Second interview	SAD
Intergenerational equity	0.37	0.24	0.13
Protection of water and other resources	0.09	0.30	0.21
Safe wastewater disposal	0.30	0.15	0.15
High social acceptance	0.17	0.23	0.06
Low costs	0.08	0.08	0.00
Total SAD			0.54

Appendix D: Decision alternatives

Table D. Main characteristics of the selected six decision alternatives, from totally 13 alternatives in Zheng et al. (subm.); also see Lienert et al. (2015). WWTP: wastewater treatment plant.

Abbreviation	Short name	Short description
A2	Central high tech	IKA (intercommunal agency) provides centralized full service WWTP, but with upgraded treatment for removal of micropollutants. New development areas outside existing building zones have separate drainage systems for rain. Rehabilitation only if length of pipes in very bad condition exceeds 10%.
A5	Decentral low tech	Wastewater is decentrally treated; responsibility of households. Cheap technologies, e.g. septic tanks and a minimal municipal collection service. No rehabilitation, only repairs upon urgent need for action.
A7	Decentral with nutrient recovery	Nearly fully decentralized system with nutrient recovery for fertilizer in agriculture; 'rest' is drained by sewer system. Cooperative and private responsibility. New areas outside existing building zones drained by separate rain water systems. Rehabilitation is realized with constant budget.
A8a	Status quo (central) with stormwater retention	Treatment of wastewater in central WWTP as today, in responsibility of communities. Additionally, max. infiltration of rain: 80% of rooftops of new buildings are connected to infiltration structures. Rehabilitation strategy as A7.
A8b	Decentral high tech, gradual transformation	In existing area, decaying WWTP; will be transformed gradually to high tech decentralized systems from 2010–2030. In new areas, high tech decentralized systems are built. Rehabilitation strategy as A7, organization as A8a.
A9	Central privatized, minimal maintenance	Water infrastructures are fully privatized, consumers choose private contractor that seek revenue maximization. Fully centralized system. Rehabilitation strategy as A5. Rest as A8a.