The Tragedy of Democratic Decision Making

Klaus Fiedler, Joscha Hofferbert, Franz Woellert, Tobias Krüger

(Universität Heidelberg)

and

Alex Koch

(University of Cologne)

To appear in:

Joseph P. Forgas, William Crano & Klaus Fiedler (Eds.),

Social Psychological Approaches to Political Psychology

Word count: xxxx

Running head: Tragedy of democracy

Author note: The research underlying the present chapter was supported by a Koselleck-grant of the Deutsche Forschungsgemeinschaft (Fi 294 / 22-2). Helpful comments were provided by Michaela Wänke on a draft of this chapter are gratefully acknowledged.

Abstract

Analogous to Hardin's tragedy of the commons – the inability to exploit the advantage of cooperation in dilemma situations – the present chapter is devoted to another major political deficit: the inability to exploit the group advantage in democratic decision making. Democracies delegate virtually all legislative, executive, and juridical decisions to groups. However, groups often fail to take advantage of the "wisdom of crowd", because leadership styles and procedural styles undermine the stochastic independence of individual opinions. The failure of groups to mobilize their potential is particularly evident in so-called hidden-profile tasks that call for effective communication of distributed information resulting from a division of labor. While previous research suggests that groups manage to find out the best decision option if only the hidden profiles are made transparent, the research reported in the present paper suggests that the failure to consider dissenters and minority arguments will persist. This relentless tragedy of democratic decision making reflects the meta-cognitive inability to suppress the impact of selective repetition on judgment formation. The discussion focuses on possible remedies that might help groups to overcome this fundamental problem.

Introduction

When social psychologists are concerned with political behavior – the common theme of the present volume – they typically focus on specifically political orientations, specific attitudes, national or group-specific identities that create the potential for conflict, harm, and discrimination. According to Wikipedia (http://en.wikipedia.org/wiki/Political_psychology), "political psychological theory and approaches have been applied in many contexts such as: leadership role; domestic and foreign policy making; behavior in ethnic violence, war and genocide; group dynamics and conflict; racist behavior; voting attitudes and motivation; voting and the role of the media; nationalism; and political extremism." The common denominator of these research topics seems to be the conflicts that arise as an inevitable consequence of different political positions, interests, and goals (Hammond, 1965).

The same emphasis of conflict proneness as a major research target is also characteristic of Garret Hardin's (1968) seminal essay "The tragedy of the commons", which illustrates the detrimental consequences of human beings' tendency to defect in social and ecological dilemma situations. Although people understand that exploiting natural resources (tropical rain forest, fish populations, water supply) or not participating in common-good institutions (health insurances, education) can cause serious social and ecological conflicts in the long run, they continue to exploit and to defect if only they believe that others exploit as well. Although Hardin's philosophical and ecological treatise does not refer to consequences of different political interests but to a general syndrome shared by all political agents, it is still concerned with the conflict proneness of defecting, uncooperative behaviors that are easy to identify and that only difficult to abandon and control: "… to avoid hard decisions many of us are tempted to propagandize for conscience and responsible parenthood. The temptation must be resisted, because an appeal to independently acting consciences selects for the disappearance of all conscience in the long run, and an increase in anxiety in the short. The only way we can preserve ad nurture other and more precious freedoms is by relinguishing the freedom to breed, and that very soon” [p. 1248].

The present chapter is concerned with another tragedy that is not so much related to conflicts created by apparent causes but, on the contrary, to a widely shared inability to detect and identify conflicts. Even when decent people do their best to cooperate in democratic settings, trying to find the best decisions in fair and impartial group discussions, they still fail to exploit the advantage of groups over individuals that is at the heart of the democratic idea. I call this failure “the tragedy of democratic decision making”, analogous to Hardin’s “tragedy of the commons”, because it casts the effectiveness of well-motivated and -reasoned democratic decisions and actions into question. Moreover, I propose that the tragedy that I have in mind originates in a syndrome what I call “meta-cognitive myopia” (Fiedler, 2012). This basic deficit is reflective of naive trust, uncritical cooperation, lack of emancipation, and the failure to recognize constructive conflicts in contrast to conflict proneness. In order to overcome this deficit, it is necessary to improve conflict-prone aspects of social intelligence, such as critical assessment, playing dissenter roles, and expressing unpopular opinions.

Group Decision Making – The Basic Democratic Paradigm

The basic method or paradigm of democratic system is the instrument of group decision making. In democratic systems, all important decisions in the service of all three powers – regarding legislative, executive and judiciary affairs – are delegated to groups. The rationale for this undisputable trust in group-like committees and parliaments does not just lie in the motive to avoid the harm caused by mad and criminal dictators or demagogues, or in the basic democratic fairness rule of one vote given to each individual. The rationale is also – and predominantly – based on the conviction that groups outperform individuals. Groups are supposed to be superior to individuals in terms of wisdom, problem-solving ability, risk assessment, accurate calculation, and in terms of political wisdom. In other words, democracies rely heavily on the wisdom of crowds (Surowiecki, 2004).

A Group Advantage Actually Exists

Increasing accuracy through aggregation. From a scientific and logical point of view, it is indeed justified to believe in a basic group advantage in judgments and decisions. In a highly uncertain world, individual estimates or predictions of political, economic or ecological developments are very error-prone. Even expert judgments bear only weak correlations to the objective criterion. A blessing feature of such a probabilistic world is that accuracy can be through error cancellation by aggregating over several independent judgments. This phenomenon is well-known from test theory. Correct responding to singular intelligence test items correlates only weakly with intelligence, but the average correctness rate across many items affords a highly reliable measure of intelligence. The Spearman-Brown formula tells us that a single-item reliability as low as r = .15 increases to over r = .80 as test length increases to 20 items. If the reliability of a one-item test is r = .25, a 10-item test reaches r = .xx. The reason for this miraculous increase in accuracy is error cancellation. By aggregating (summing, averaging) over many items, the error component of individual items is canceled out. Some items over-estimate and other under-estimate the test person’s ability, so that the overall error component is levelled off, whereas the systematic variance component shared by all items (viz., intelligence) becomes stronger and stronger.

The same aggregation advantage, which can be predicted on safe mathematical grounds, applies to group performance. Even when individual group members’ judgments are only weakly correlated with a correctness criterion, an aggregate (average) score computed from an increasing number of group members must increasingly reflect the systematic knowledge shared by all judges while abstracting from individual judges’ error. The same Spearman-Brown formula (or related methods suggested by Rosenthal, 1987) can be used to quantify the expected accuracy gain. As a matter of principle, then, there is a rational basis for the democratic paradigm; groups do have the potential to outperform individuals due to a well-understood aggregation effect. This uncontested performance advantage that results from averaging two or more indicators is the origin of the proverbial wisdom of crowd – and a fundamental asset of all democracy.

The failure to exploit the aggregation advantage. Unfortunately, however, some conditions must be met to exploit the depicted group advantage and human beings – even the smartest and most educated ones – fail to establish or sometimes even avoid these conditions. The most important precondition for profiting from an aggregation advantage is stochastic independence. Just test reliability only increases with test length when the responses to different items are independent, groups can only profit from the wisdom of an more than one person if different group members' knowledge and wisdom reflects independent sources and can be uttered freely, unrestricted by conformity and desirability constraints. However, for various reasons, existing group and group leaders do everything to undermine this principle of independence. Group members often receive their information from the same or from overlapping sources, rendering them all subject to the same systematic biases. Dominant leaders force group members to follow their preferred opinions and to adopt their privileged information, and dissenters are sanctioned and discouraged to freely present deviant arguments (Schulz-Hardt, Brodbeck, Mojzisch, Kerschreiter & Frey, 2006). Unanimity rules (Kameda & Sugimori, 1993) and time pressure restrict group discussions to one or very few mainstream themes, leading to group polarization (Myers & Lamm, 1976) and sometimes almost pathological group-think effects (Baron, 2005; Janis, 1972). As a consequence, group discussions are characterized by a marked shared-information bias (Larson, Foster-Fishman & Keys, 1994), that is, a preference to contribute redundant but desirable arguments that are already shared by others in the group, rather than uttering new and potentially deviant arguments.

Such a shared information effect, which clearly undermines stochastic independence in group interaction, has been attributed to social motives, such as need for approval, the goal to get along with others (Hollingshead, Jacobsohn & Beck, 2007), or to a lack of self-confidence and courage. However, although such motives may clearly contribute to the streamlining and conformity syndrome that undermines democratic decision processes, they can hardly account for the entire problem. Even in the absence of prosocial motives or cowardice, or when group pressure is fully absent, individuals will succumb to the same mistakes that prevent them from exploiting aggregation effects. Pertinent evidence come from recent research on advice taking (Yaniv, Choshen-Hillel & Milyavsky, 2009) – a paradigm in which individual judges or decision makers (rather than groups) try to exploit the wisdom of advice givers. When provided with estimates of one or two advisors, in addition with their own estimates, individual judges do not base their ultimate judgments or decisions on an aggregate of all independent estimates. Rather, they typically give more weight to their own estimate (reflecting an egocentric bias), and they give more weight to those advisors that share their own perspective (reflecting a redundancy bias). When given a choice between diverse advisors, they prefer to ask those whose judgments are consistent with their own. This preference for confirmatory and redundant advice, which cannot be attributed to expected validity, is the opposite of what normative models of knowledge aggregation prescribe (Larrick & Soll, 2006). According to these normative considerations, the expected information gain is maximized when the combined sources or knowledge are as independent as possible.

***Metacognitive Myopia – A Severe Drawback of Rational Decision Making*.**

Thus, the failure to exploit the aggregation effect in group decision making is neither specific to actual group settings nor to group-related motives and behaviors. Rather, the syndrome is anchored in a more fundamental deficit of homo sapiens that I have come to call meta-cognitive myopia (Fiedler, 2012). What prevents people from rational decision making, whether in groups or individually, is not so much the truth heuristic (Dechêne, Stahl, Hansen & Wänke, 2010) that lets one’s own familiar opinions or the opinions shared with others appear socially validated and true. Following such a social heuristic (Hoffrage & Hertwig,

2012) could actually be an asset in an uncertain world, provided some metacognitive monitoring and control system prevents individuals from overstretching its value. What has to be blamed as a most serious obstacle in rational and responsible decision making is the naïve and uncritical manner in which people fail to correct for obvious mistakes and drawbacks of their intuitions, gut feeling, and social heuristics. Because this notorious failure to engage in a critical evaluation of the differential value of information samples reflects a deficit in monitoring and control functions, it is called metacognitive myopia.

The analogy to “myopia” (short-sightedness) highlights the fact that people are often remarkably accurate in processing the information given (e.g., in a group discussion or data sample), even on cognitively demanding tasks. However, despite this basic sensitivity to the information given in a sample, they fail to engage in a critical examination of the validity and usability of information samples, even when they are obviously and blatantly biased. Nevertheless, people will continue to use and (accurately) process the given information, uncritically and naively (Juslin, Winman & Hansson, 2007; Stewart, Chater & Brown, 2006), rather than separating the wheat from the chaff and only utilize samples that promise to be unbiased and not misleading.

Let us consider a few illustrative examples of metacognitive myopia before we return to research on group decision making. In Jones and Harris’ (1967) seminal demonstration of the fundamental attribution error, participants inferred an essay writer’s attitude (pro vs. contra Fidel Castro) even when they had been told that the writer had been assigned a position to be advocated in the essay. Still, the information given in the text sample was readily used to infer the author’s attitude although it was obviously not valid. In a similar vein, probability and risk judgments exhibit a strong and regular denominator neglect (Reyna & Brainerd, 2008), taking the absolute number of critical outcomes for granted and not taking the overall size of a sample for granted. For instance, teachers fail to divide their count of different students’ correct answers by the respective sample size (i.e., the number of responses provided by different students; Fiedler & Walther, 2004). In one particularly consequential and politically relevant variant of this denominator neglect, legal expert witnesses count the number of linguistic truth signs in the witness' report without normalizing the count for text length (Vrij, 2005). Consumers and Internet users believe that the top entries in a googled list provides a representative picture of reality much like scientists proceed as if the published literature provides a valid picture of the empirical reality.

Such obvious violations of the need to correct for sample biases are particularly noteworthy in situations in which individuals themselves have obscured their information samples in their own lop-sided information search process. For example, in a series of studies by Fiedler, Brinkmann, Betsch and Wild (2000), participants were asked to take the perspective of politically mature citizens whose task is to critically assess health related risks. One problem called for an assessment of the probability *p*(breast cancer | positive mammogram) that a woman has breast cancer given a positive mammogram. Participants could themselves draw a sample of relevant information from an index card file. Each index card in the file contained information about a woman's mammogram on one side (positive or negative) and her diagnosis on the other side (breast cancer vs. no breast cancer). The distribution of cards resembled the real distribution. In reality, the baserate of women who have breast cancer is about 1% (i.e., 10 out of 1000), the hit rate of women with breast cancer who are tested positively is about 80% (i.e., 8 out of 10) and the false alarm rate of positively tested women without breast cancer is somewhat less than 10% (i.e., 99 out of 990). Note that from the absolute frequencies presented in parentheses, it is clear that of all 107 women who are tested positively, by far the greatest part does not have breast cancer. The correct estimate of *p*(breast cancer | positive mammogram) is therefore less than 10%.

In one experimental condition, participants sampled breast cancer data conditional on the mammogram results. That is, they could draw as many cards as they wanted from an index-card file with two slots, one for positively tested and one for negatively tested women.

In this sampling condition, which is logically appropriate to assessing *p*(breast cancer | positive mammogram), most participants understood that they only have to consider the slot for women with positive test results, and they easily found out that the posterior probability of breast cancer given a positive mammogram is rater low. However, in another condition, the file contained two slots for women with and without breast cancer, and they could sample mammogram data conditional on diagnosis. In this condition, they would typically draw all those few cases with breast cancer plus approximate the same number of women without breast cancer. Thus, although they could clearly see the small baserate of breast cancer in the entire file, they sampled roughly the same number of cases from both slots, thereby drastically exaggerating the breast cancer rate in the sample. Nevertheless, this did not prevent them from estimating *p*(breast cancer | positive mammogram) from the respective proportion in the sample, although they had themselves determined it to be roughly 50%. Note that the serious over-estimations obtained in this experimental group came along with high accuracy. The estimate closely corresponded to the sampled proportions *p**(breast cancer | positive mammogram), and they did not confuse this task-relevant proportion with the reverse proportion *p**(positive mammogram | breast cancer). Despite this basic sensitivity to the sample given, their meta-cognitive myopia prevented them from recognizing the highly misleading composition of samples that grossly over-represented the breast cancer rates.

New Empirical Evidence for the Tragedy of Democracy

The remainder of this chapter is devoted to reviewing some empirical studies that speak to the role of metacognitive myopia in group decision making and in politically relevant problem contexts. On one hand, virtually all pertinent prior research (Fiedler, 2000, 2008a, 2012; Moore & Healy, 2008) has been confined to individual judgment and reasoning. Finding similar evidence from group-decision tasks is therefore of immediate interest to political psychology. On the other hand, one might conjecture that conflict-prone group settings prime more critical, less naïve reasoning so that groups might be more likely to overcome the limitations of metacognitive myopia than individuals. Let us first pursue this possibility with regard to a similar conditional-sampling task as in the aforementioned studies by Fiedler et al. (2000).

Conditional Sampling in Dyads: Myopia for Useless and Misleading Samples

In a series of recent experiments (Fiedler, Krüger & Koch, 2014), two people working together in dyadic decision tasks were exposed to conflicting samples of information. They were asked to jointly determine the probability of winning in different lotteries or the probability that departments in an organization win a price after they have either participated in a training program or not. In all experiments, the decision task called for the comparison of the winning rates, p(win | A) and p(win | B) of two options, A and B. For convenience, let A always refer to the superior option with a slightly but noticeably higher winning rate. One person in the dyad, the valid sampler, could sample the correct way, gathering observations about winning conditional on prior examples drawn from A or B. For example, valid sampling would reveal winning rates of p(win | A) = .60 and p(win | B) = .40 for options A and B, respectively. Thus, apart from sampling error and memory loss, valid samplers should correctly recognize the superiority of option A over B. The other person in the dyad, however, was assigned the role of an invalid sampler, being exposed to conditional samples of A versus B conditional on winning or not winning outcomes. Although drawn from the same options (A vs. B) x outcomes (win vs. not win) distribution as valid samples, invalid samples could produce quite different associations between winning and the two options A and B. For example, in one experiment, the probability p(A | win) that a given winning trial was associated with A (the superior option) could be as different as .27 and .75 in different conditions, depending on the baserate of A (versus B) in the data base. If the A baserate is very low (i.e., if there are much fewer A than B cases), the probability of A given a winning outcome may be very low even when A is the superior option. Thus, while invalid samples were always logically useless – because p(A | win) must not be confused with p(win | A) –

they happened either to support a correct decision (when A was frequently associated with winning) or to interfere with a correct decision (when A was frequently associated with winning in the sample).

Both participants in a dyad first gathered their own sample separately and made an individual estimate of p(win | A), the task-relevant quantity (, which was, notably, the reverse conditional of what invalid samplers had observed). They were then asked to form a final estimate in a dyadic group discussion. To the extent that they critically discussed what information they had at their disposal, the dyads should have figured out that only valid samplers' evidence can be used and that invalid samplers' evidence is useless or even severely misleading. Recognizing this difference in validity and usability of samples should be particularly motivated in conflict-prone dyads, in which the valid and invalid samplers formed highly discrepant associations between A and winning.

However, the results of all three experiments showed that although dyads did agree that valid samplers possessed more useful information than invalid samplers, the overall result reflected a shady compromise. Group judgments reflected a weighted average of both group members, with only slightly higher weight given to the valid sampler but sufficient weight given to the invalid sampler that the final judgments were seriously biased toward the output-bound proportion p(A | win), which is categorically different from p(win | A). Because the invalid quantity p(A | win) varied between experimental conditions, whereas the valid quantity p(win | A) remained constant, the irrelevant variation in the former was the chief determinant of decision correctness across dyads. Subjective confidence was largely unaffected by the fact that a blatantly biased sample had intruded into the group discussion.

It is important to note that in this experimental setup, overcoming myopia means not to aggregate (average) both judgments but to discriminate between a valid and an invalid judgment that ought to be discarded. Appropriate metacognitive monitoring and control would amount to selective processing. However, almost all dyads act as if there is a social rule to do justice to every group member's sample, regardless of its validity. It is as if aggregating over both samples, however unequal and incompatible they are, is supposed to be "fair" or "natural". However, again, such a seemingly plausible social motive does not make much sense, because both members in a dyad could improve their performance from jointly recognizing that only one sample is valid. Even extra monetary rewards announced for correct joint decisions did not improve the performance. Moreover, there was no tendency for the most conflicting constellations (i.e., discrepancies between valid and invalid sample implications) to instigate better groups reasoning and decision making.

Reasoning errors due to inadequate conditional sampling constitute a huge problem for many political and economic decision problems. For instance, when politicians or legal experts estimate the causal impact of alcohol consumption on the occurrence of traffic accidents, the logically appropriate conditional *p*(accident | alcohol consumed), the probability of accidents given the consumption of (some specified quantity of) alcohol. However, it is legally, ethically, and pragmatically impossible to assess this probability, because one cannot let people consume alcohol and then wait (for months or years) to see how many accidents happen. In such cases, it is common place to use the reverse conditional, *p*(alcohol consumed | accident), as a surrogate. This conditional probability is easily available for whenever an accident happens, one can assess if the driver has consumed alcohol. To be sure, *p*(alcohol consumed | accident) is many times higher than *p*(accident | alcohol consumed) and useless as an estimate of the causal impact of alcohol on traffic accidents.

In a similar vein, after the 09/11 terror attack, many organizations prohibited their managers from using airplanes, apparently because the association between air traffic and lethal outcomes was based on the only available conditional, *p*(using airplane | lethal outcome). The reverse conditional, *p*(lethal outcome | using airplane), which is logically relevant to judging the danger of flying in an airplane, could not be assessed (cf. Gavanski & Hui, 1992). As a consequence of this category mistake, the rate of lethal traffic accidents on the road, due to people forces to drive far distance on the road rather than taking an airplane, exceeded the rate of lethal 09/11 victims within only three months (i.e., before the end of the year 2001; Gigerenzer, 2004). Apparently, the confusion of reverse conditional probabilities can have detrimental consequences.

Given such pitfalls in conditional reasoning tasks, it would appear to be essential that democratic decision groups not just establish fair majority rules and cancel off error through aggregation. Rather, in addition to exploiting the wisdom of crowds, one would expect democratic groups also to overcome or ameliorate the meta-cognitive myopia that prevents individuals from conditional reasoning. Collective intelligence should prevent deliberating groups from using apparently biased and inadequate samples uncritically, when conflicting individual knowledge instigates critical discussion and more sophisticated reasoning (Mata, Fiedler, Ferreira & Almeida, 2013). However, as evident from the Fiedler et al. (2014) studies, no support for this optimistic expectation was found for dyads. On the contrary, research on group polarization (Forgas, 1990; Myers & Lamm, 1976) and groupthink (Baron, 2005) suggests that group discussion may further strengthen initially existing biases. Rather than being sensitized by conflicting and dissenting opinions, group discussions typically serve to reinforcing the dominant majority. The resulting amplification of the most widely shared majority positions then often amounts to worsening rather than overcoming the myopious trust in biased samples. Rich evidence for this memorable phenomenon comes from controlled experiments using hidden profiles.

Hidden Profiles: Myopia for the Impact of Repetition on Cognitive Inferences

The hidden-profile paradigm (Mojzisch & Schulz-Hardt, 2006; Stasser & Titus, 1985) is the “drosophila” for researchers concerned with group-decision failure. A hidden-profile is a configuration of advantages and disadvantages of several decision options (A, B, C, etc. such as candidates in a personnel-selection task or action strategies in a planning task) that is unknown to individual group members. So the distributed knowledge has to be uncovered through effective communication. When there is a division of labor such that every group member has gathered only partial information, it is possible that although A is actually the best option, B may appear to be better from the subset of information available to all individual group members. This may occur, for instance, when A's few disadvantages are shared (i.e., fully known to all members) whereas A's many advantages are unshared such that each of, say, four group member knows only one fourth. In contrast, the many disadvantages of the worse option B may be unshared (i.e., distributed over different group members) whereas the few advantages may be shared. Given such a constellation, the best alternative can only be identified if group members do not adhere to their premature, pre-discussion preferences but manage to make their collective knowledge (i.e., the hidden profile) transparent through effective group discussion.

Numerous empirical studies testify to the inability of even highly motivated groups to solve such hidden-profile problems (cf. Kerr & Tindale, 2004; Mojzisch & Schulz-Hardt, 2006). This persistent failure is often due to the reluctance to discuss unshared information, either because unshared arguments are socially less satisfying than shared arguments, or because they are appear less valid or inconsistent with pre-discussion preferences (cf. Mojzisch & Schulz-Hardt, 2006). Another reason why groups cannot solve hidden profiles is because they do not engage in any discussion of raw arguments but confine themselves to negotiating the individual preferences. For instance, applying a majority rule will lead to a wrong decision if all or most members hold mistaken preferences.

In any case, whatever causes or reasons prevent groups from considering unshared information, it is generally presupposed that groups can easily understand and solve the decision problem as soon as the hidden profile is made fully transparent. Once the positive and negative aspects of all decision options are presented openly, the dominant option should be apparent. However, recent research inspired by the notion of metacognitive myopia shows that the difficulty with hidden profile tasks will persist even when all information is revealed openly. In a series of experiments using a typical personnel selection task (Fiedler, Hofferbert & Woellert, 2014), participants received tape-recorded descriptions of positive and negative attributes of four choice candidates, A, B, C, and D, provided by other members of a fictitious groups. Both the differences and ratios of the number of positive and negative attributes revealed a clear-cut preference order of the candidates, A < B < C < D, which was easily identified in a control condition. In an experimental condition, the same fully transparent information was provided but some attribute descriptions were repeated more often than others, as can be expected in a group discussion when some pieces of information are shared more than others. Through selective repetition of the candidates' attributes, the clearly inferior candidate B's advantages and the best candidate D's disadvantages were over-represented in the group report. Thus, although an actual count of the *types* of candidate attributes clearly revealed the actual ordering, A < B < C < D, a count of the *tokens* suggested a misleading ordering, B < A < D < B.

Note that this manipulation of selective repetition is tantamount to the way in which majority and minority arguments are represented in group discussion (Gerard, 1985). If the same discussion time and opportunity is given to majority and minority members, the arguments shared by majorities can be naturally and inevitably expected to be repeated more frequently than the arguments held by the minorities. Thus, even when all arguments are made perfectly transparent, selective repetition will give a learning advantage of majority arguments, simply because learning is a monotonous function of the number of repetitions or learning trials. Ignoring this natural learning advantage of majority arguments thus constitutes a common variant of metacognitive myopia. Conversely, for democratic groups to overcome such myopia amounts, they have to monitor and control for the impact of selective repetition in an attempt not to be led or misled by the superficial repetition of completely redundant information. Note that this ambitious meta-cognitive task means to counteract an unwarranted wisdom-of-crowd effect, in a deceptive situation in which "the crowd" does not consist of is independent voices but only reflects the fully redundant and uneven repetition of pro arguments for one option and contra arguments against another option.

To establish such a distorted pattern in the experiments conducted by Fiedler et al. (2014), the participants were instructed that part of the information would be repeated for technical reasons (i.e., due to suboptimal methods of cutting the audio-tape used to convey the group members' reports). In different experiments, repetitions came from the same voice rather than from a different group member to rule out the possibility that repeated arguments might be socially validated. Participants in one condition were explicitly warned not to be misled by mere repetitions, which could bias their judgments. However, all attempts to undo the impact of repetition on attitude learning were in vain. Candidate B, who profited most from repetition of majority information, was consistently preferred to the actually superior Candidate D, who suffered most from repetition of majority information.

From a perspective of a metacognition researcher, this inability not to learn from repetition makes perfect sense. Just as we cannot tell out perceptual system to see a snake or a cold drink if these stimuli are obviously there, we cannot tell our brain not to acquire stronger memory representation from increasing numbers of learning trials (Unkelbach, Fiedler & Freytag, 2007). Moreover, people are metacognitively blind for this inevitable advantage of repeated information. This is nicely illustrated in a study by Koriat (1997), in which participants who had been presented with word pairs like *bread – cheese* and had to estimate the probability that they can later recall one word given the other word as a cue. When some word pairs were repeated, participants complained that they had already indicated the probability before, thus making the new judgment superfluous. There was little insight into the obvious fact that recall probabilities should increase across repeated presentations of the same stimulus pairs.

To be sure, although participants in the Fiedler et al. (2014) were unable to suppress the encoding and memorization of repeatedly presented arguments during the learning stage, they were later able to correct their judgments if they were explicitly told repetitions had created an unwarranted disadvantage of D and advantage of B. Given such an instruction, they were of course able to downgrade B and to upgrade D on the final ratings scales. However, this superficial correction effect must not be mistaken for mastery of the tragedy of democratic decision making. Although they could follow the demand and correct their local judgments, when the same participants were later told to retell other people what they knew about candidates A, B, C, and D, the repetition bias was still alive. Regardless of their corrected judgments, they could recall more repeated information and, if anything, participants who had been asked to suppress their intuitive ratings exhibited a slightly stronger repetition bias than participants who had corrected themselves. Thus, the repetition bias had become social reality, notwithstanding the possibility to correct for local judgments.

It should be noted in passing that the same repetition bias that gives a learning advantage to arguments provided *by* majorities (rather than minorities) also reappears in learning *about* majority versus minority behaviors (Fiedler, 1996). When the same high rate of (say 75%) positive behaviors is observed in a majority (i.e., in a large sample) and in a minority (i.e., in a small sample), the positivity trend will be learned more completely for the larger sample, as already explained before. Various independent experiments and computer simulation studies have shown that such an illusory correlation (Hamilton & Gifford, 1976) bias against minorities can arise in the absence of any biased attention or memory, simply as a reflection of learning curves increasing with number of trials (Fiedler, 1996, 2000; Kutzner, Vogel, Freytag & Fiedler, 2011; Smith, 1991).

The meta-cognitive failure to monitor and control for such an advantage of majorities or large samples has also been called denominator neglect (Reyna & Brainerd, 2008). When 24 positive behaviors in a sample of 30 is worth more than 12 in a sample of 15, this reflects the concentration on the absolute frequencies in the numerator (24 > 12) while ignoring the equally different set sizes in the denominator of the ratios 24/30 and > 12/15. Democratic institutions and committees make little attempts to correct for this denominator neglect. When evaluating the success of organizations (like university departments or firms), a common index is the number of publications or the gross income. It is quite uncommon to divide this numerator by the denominator reflecting the set size or input into the evaluated unit. When a publication in an interdisciplinary journal like Science or Nature is commonly considered most excellent, evaluation committees would hardly divide the journal's high citation index of the journal by the huge set size of the readership. Or, to provide an example from judiciary decision making, when existential courtroom decisions have to rely on an expert's analysis of a witness' credibility using criteria-based content analysis (Vrij, 2005), in the absence of physical evidence, the number of linguistic truth criteria found in the witness report is not normalized for text length. Politicians, executive officers, journalists, and many other agents in democratic groups and organizations do not seem to have the slightest interest in changing the pervasive denominator neglect.

Apparently, then, the treatment of metacognitive myopia cannot be reduced to instructions for group leaders to exploit the wisdom of crowds, to allow dissenters to utter their opinion and to render hidden profiles transparent. Even when all these criteria of good democratic discussion style are met, there is still the differential impact of majorities and minorities on the cognitive process of learning and inference making. Coping with this problem requires groups to forego the wisdom-of-crowd effect; they must not follow an information-aggregation process if knowledge items are not independent, representing merely redundant repetitions of already given information. However, classifying arguments as either old and redundant with previous statements or as new and original can be a highly difficult judgment task in itself, especially when group members are exposed to the same media and information sources. Eventually, the longer one reflects on ways of the mastering the tragedy of democratic decision making, the more monstrous and hard to resolve the tragedy appears.

Aggregation Levels: Myopia for Different Truths Holding at the Same Time

All the manifestations of metacognitive myopia that we have considered so far are presumably easier to handle than the upsetting fact that different truths exist at the same time. As a consequence, different political interest groups can justify their divergent convictions with reference to different “facts” existing at different levels of aggregation. Just as drinking alcohol can be arguably said to be enjoyable today, causing a hang-over tomorrow, the key to good social contact in the next weeks, and serious liver disease over years, the same nations may be considered rich (in terms of the national gross product) but also poor (in terms of high prevalence of poverty). In a dilemma game, defecting maximizes the payoff at the level of single trials, whereas cooperative strategies are more successful when payoffs are aggregated across many trials (Fiedler, 2008b). More generally, delay of gratification tasks highlight the fact that what is pleasant in the short run may be detrimental in the long run and vice versa (Metcalfe & Mischel, 1999).

The mastery of meta-cognitive myopia vis-à-vis such challenging task situations calls for an ability to accept two or more valid truths simultaneously. In political settings, it is almost impossible to communicate and negotiate solutions with more than one optimum (Coombs & Avrunin, 1977). Solving such problems is commonly considered politically not feasible. Research on Simpson’s paradox (Fiedler, Walther, Freytag & Nickel, 2003) and on ecological-correlations shows that it is almost impossible to induce an understanding of the fact that, say, more male than female students are accepted for graduate studies (suggesting discrimination against females) although the rate of accepted females is higher within both graduate programs offered by the same university. This paradox is possible when most women apply for the more difficult graduate program with a generally much higher rejection rate than an easier graduate program where most males apply. Although people do understand similar constellations – for instance, that a better tennis player can have a lower winning record in a higher league than an inferior player in a lower league – participants have a very hard time understanding that two (or more) truths can hold at the same time.

Concluding Remarks

Thus, it appears as if the tragedy of democratic decision making cannot be reduced to a few stupid habits and easily correctable mistakes. Rational decision making in democratic groups takes more than exploiting the wisdom of crowds and majorities. Simply aggregating over many opinions, asking many advisors, applying majority or Condorcet rules (Hastie & Kameda, 2005) or revealing the full profile of all available information is not enough to exploit the potential of democratic groups. Rather than merely aggregating over large samples of observations and agents, responsible democratic decisions are first of all contingent on a metacognitive understanding of the problem structure (Mercier & Landemore, 2012). The wisdom of crowds can only reduce unsystematic error; it is of little help when the group members' information samples are subject to serious bias, due to inappropriate sampling procedures. Pooling biased information samples may worsen rather than improving the validity of decision. Complex data patterns convey different messages at different levels of analysis, and the resulting conflicts cannot be cancelled out by averaging information obtained at different levels.

Therefore, the first and foremost maxim of rational decision making is not to maximize and exploit the quantity of information but to monitor and control the quality of information. The most prominent task for democratic institutions is to create a climate for critical assessment and emancipation that increases the likelihood of detecting and circumventing the various pitfalls in the sampling and discussion process that we have discussed in the preceding sections. In such a climate, the tragedy of democracies may be alleviated and sometimes circumvented when mature groups engage in three stages of metacognitive deliberation. In the first stage, it is useful to identify those straightforward situations in which information samples can be assumed to be unbiased and unproblematic. It is these task settings that promise to profit from the wisdom of crowds, from advise taking and aggregation of independent opinions. If however information sampling is potentially biased, the key to rational decision making is no longer the group’s ability to aggregate many opinions and large data arrays but to discriminate between useful (unbiased, adequate) and useless (biased, inadequate) samples. In these situations, any compromise that tries to do “justice” to each and every opinion must be discarded as shady and dangerous. Compromises in these situations do not reflect “realistic” of “feasible” policy making but maybe the worst type of metacogitive myopia. In a third and final stage, then, it is essential to reflect on the multidimensional nature of most utility functions. An open-minded utility analysis will often reveal that different aggregation levels, time windows, and attribute weighting will yield different orderings of the best decision options that are finally retained in a short list.

To be sure, some aspects of the tragedy can never be fully mastered, such as the inevitable learning and memory advantage of majority opinions. The only remedy would be to artificially allocate more reiteration time for minority opinions, but such a procedure would not be fair and democratic either. Similarly, memory encoding and organization will often support the extraction of more superordinate differences at high aggregation level than the extraction of more specific and varied low-level regularities (Fiedler, Freytag & Meiser, 2009). It is unlikely that all metacognitive obstacles will be removed. However, I hope and I actually believe that the quality of democratic decisions can be improved to the extent that groups try to be not just useful crowds but collective institutions that encourage metacognitive monitoring and control processes.

References

Baron, R. S. (2005). So Right It's Wrong: Groupthink and the Ubiquitous Nature of Polarized Group Decision Making. In M. P. Zanna (Ed.) , *Advances in experimental social psychology, Vol. 37* (pp. 219-253). San Diego, CA US: Elsevier Academic Press.

Coombs, C. H., & Avrunin, G. S. (1977). Single-peaked functions and the theory of preference. *Psychological Review*, *84*(2), 216-230. doi:10.1037/0033-295X.84.2.216

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality And Social Psychology Review*, *14*(2), 238-257. doi:10.1177/1088868309352251

Fiedler, K. (1996). Explaining and simulating judgment biases as an aggregation phenomenon in probabilistic, multiple-cue environments. *Psychological Review*, *103*(1), 193-214. doi:10.1037/0033-295X.103.1.193

Fiedler, K. (2000). Beware of samples: A cognitive-ecological sampling approach to judgment bias. *Psychological Review*, *107*, 659– 676.

Fiedler, K. (2008a). The ultimate sampling dilemma in experience-based decision making. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *34*(1), 186-203. doi:10.1037/0278-7393.34.1.186

Fiedler, K. (2008b). Pseudocontingencies – A key paradigm for understanding adaptive cognition. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for rational models of cognition* (pp. 165-186). London: Oxford University Press.

Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical inference. In B. H. Ross (Ed.) , *The psychology of learning and motivation (Vol 57)* (pp. 1-55). San Diego, CA US: Elsevier Academic Press.

Fiedler, K., Brinkmann, B., Betsch, T., & Wild, B. (2000). A sampling approach to biases in conditional probability judgments: Beyond base rate neglect and statistical format.

Journal of Experimental Psychology: General, *129*(3), 399-418. doi:10.1037/0096-3445.129.3.399

Fiedler, K., Freytag, P., & Meiser, T. (2009). Pseudocontingencies: An integrative account of an intriguing cognitive illusion. *Psychological Review, 116*(1), 187-206. doi:10.1037/a0014480

Fiedler, K., Hofferbert, J., & Woellert, F. (2014). *Sharing information is not enough: Repetition biases in hidden-profile situations*. Manuscript submitted for publication.

Fiedler, K., Krüger, T., & Koch, A. (2014). *Dyadic judgments based on conflicting samples: The inability to ignore invalid input*. Manuscript submitted for publication.

Fiedler, K., Walther, E., Freytag, P., & Nickel, S. (2003). Inductive reasoning and judgment interference: Experiments on Simpson's paradox. *Personality and Social Psychology Bulletin*, *29*(1), 14-27. doi:10.1177/0146167202238368

Forgas, J. P. (1990). Affective influences on individual and group judgments. *European Journal of Social Psychology*, *20*(5), 441-453. doi:10.1002/ejsp.2420200506

Gavanski, I., & Hui, C. (1992). Natural sample spaces and uncertain belief. *Journal of Personality and Social Psychology*, *63*(5), 766-780.

Gerard, H. B. (1985). When and how the minority prevails. In S. Moscovici, G. Mugny, & E. Van Avermaet (Eds.), *Perspectives on minority influence* (pp. 171-186). Cambridge: Cambridge University Press.

Gigerenzer, G. (2004). Dread risk, September 11, and fatal traffic accidents. *Psychological Science*, *15*(4), 286-287. doi:10.1111/j.0956-7976.2004.00668.x

Hamilton, D., & Gifford, R. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*(4), 392-407. doi:dx.doi.org/10.1016/S0022-1031(76)80006-6

Hammond, K. R. (1965). New directions in research on conflict resolution. *Journal of Social Issues*, *21*(3), 44-66.

Hardin, G. (1968). *The tragedy of the commons*. *Science, New Series,162*, No. 3859, 1243-1248.

Hastie, R., & Kameda, T. (2005). The robust beauty of majority rules in group decisions. *Psychological Review, 112*, 494–508.

Hoffrage, U., & Hertwig, R. (2012). Simple heuristics in a complex social world. In J. I. Krueger (Ed.) , *Social judgment and decision making* (pp. 135-150). New York, NY US: Psychology Press.

Hogarth, R. M. (2001). *Educating intuition*. Chicago: University of Chicago Press.

Hollingshead, A. B., Jacobsohn, G., & Beck, S. J. (2007). Motives and Goals in Context: A Strategic Analysis of Information Sharing in Groups. In K. Fiedler (Ed.) , *Social Communication* (pp. 257-280). New York, NY US: Psychology Press.

Jones, E. E., & Harris, V. A. (1967). The Attribution of Attitudes. *Journal of Experimental Social Psychology*, *3*(1), 1-24. doi:10.1016/0022-1031(67)90034-0

Juslin, P., Winman, A., & Hansson, P. (2007). The naïve intuitive statistician: A naïve sampling model of intuitive confidence intervals. *Psychological Review*, *114*(3), 678-703. doi:10.1037/0033-295X.114.3.678

Kameda, T., & Sugimori, S. (1993). Psychological entrapment in group decision making: An assigned decision rule and a groupthink phenomenon. *Journal of Personality and Social Psychology*, *65*(2), 282-292. doi:10.1037/0022-3514.65.2.282

Kerr, N. L., & Tindale, R. (2004). Group performance and decision making. *Annual Review of Psychology*, *55*623-655. doi:10.1146/annurev.psych.55.090902.142009

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349-370. doi:10.1037/0096-3445.126.4.349

Kutzner, F., Vogel, T., Freytag, P., & Fiedler, K. (2011). A robust classic: Illusory correlations are maintained under extended operant learning. *Experimental Psychology, 58*(6), 443-453. doi:10.1027/1618-3169/a000112

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the avergaging principle. *Management Science*, *52*(1), 111-127. doi:10.1287/mnsc.1050.0459

Larson, J. R., Foster-Fishman, P. G., & Keys, C. B. (1994). Discussion of shared and unshared information in decision-making groups. *Journal of Personality and Social Psychology*, *67*(3), 446-461. doi:10.1037/0022-3514.67.3.446

Mata, A., Fiedler, K., Ferreira, M. B., & Almeida, T. (2013). Reasoning about others' reasoning. *Journal of Experimental Social Psychology*, *49*, 486–491.

Mercier, H., & Landemore, H. (2012). Reasoning is for arguing: Understanding the successes and failures of deliberation. *Political Psychology*, *33*(2), 243-258. doi:10.1111/j.1467-9221.2012.00873.x

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review*, *106*(1), 3-19. doi:10.1037/0033-295X.106.1.3

Mojzisch, A., & Schulz-Hardt, S. (2006). Information Sampling in Group Decision Making: Sampling Biases and Their Consequences. In K. Fiedler, P. Juslin, K. Fiedler, P. Juslin (Eds.) , *Information sampling and adaptive cognition* (pp. 299-326). New York, NY US: Cambridge University Press.

Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83*(4), 602-627. doi:10.1037/0033-2909.83.4.602

Rosenthal, R. (1987). *Judgment studies: Design, analysis, and meta-analysis*. New York, NY Paris USFrance: Cambridge University Press.

Schulz-Hardt, S., Brodbeck, F., Mojzisch, A., Kerschreiter, R., & Frey, D. (2006). Group decision making in hidden profile situations: Dissent as a facilitator for decision quality. *Journal of Personality and Social Psychology, 91*(6), 1080-1093.

Smith, E. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology*, *27*(2), 107-123. doi:10.1016/0022-1031(91)90017-Z

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, *48*(6), 1467-1478. doi:10.1037/0022-3514.48.6.1467

Stewart, N., Chater, N., & Brown, G. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1-26. doi:dx.doi.org/10.1016/j.cogpsych.2005.10.003

Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. New York, NY US: Doubleday & Co

Unkelbach, C., Fiedler, K., & Freytag, P. (2007). Information repetition in evaluative judgments: Easy to monitor, hard to control. *Organizational Behavior and Human Decision Processes*, *103*(1), 37-52. doi:10.1016/j.obhdp.2006.12.002

Vrij, A. (2005). Criteria-Based Content Analysis: A Qualitative Review of the First 37 Studies. *Psychology, Public Policy, And Law, 11*(1), 3-41. doi:10.1037/1076-8971.11.1.3

Wittenbaum, G. M., & Park, E. S. (2001). The collective preference for shared information. *Current Directions in Psychological Science*, *10*(2), 70-73. doi:10.1111/1467-8721.00118

Yaniv, I., Choshen-Hillel, S., & Milyavsky, M. (2009). Spurious consensus and opinion revision: Why might people be more confident in their less accurate judgments?. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *35*(2), 558-563. doi:10.1037/a0014589