Metacognitive Myopia – Gullibility as a Major Obstacle in the Way of Irrational Behavior

Klaus Fiedler

(University of Heidelberg)

Running head: Gullibility and irrationality

Author Note: Email correspondence may be addressed to kf@psychologie.uni-heidelberg.de

Abstract

The term “meta-cognitive myopia” (MM) refers to the phenomenon that people are pretty accurate in processing even complex information, but they are uncritical and naïve regarding the history and validity of the information given. This naive reliance on given evidence (hearsay, social-media tweet, gossip, advertising, others’ opinions) is most conspicuous when the task context makes it crystal-clear that the evidence is biased and should not be trusted. Yet, judgments and decisions are nevertheless influenced by such invalid evidence. In this article, I illustrate MM in various paradigms: inability to ignore irrelevant information; change detection; conditional reasoning; judgment of causal power; and myopia for the impact of aggregation levels. MM offers alternative accounts of many prominent anomalies and biases in judgment and decision making, but also novel predictions of formerly unknown anomalies. A gullibility perspective on MM highlights the social responsibility to monitor and control rational behavior at the metacognitive level.

Introduction

For more than half a century, psychological research has been concerned with unwanted consequences and serious costs of irrational judgments and decisions. This provocative research topic emerged in the late sixties and in the early seventies of the last century, shortly after a rationalist view on homo sapiens had been established in developmental research (Piaget, 1950), reasoning (Sarbin, Taft & Bailey, 1960), and in the social psychology of attribution (Jones & McGillis, 1976; Kelley, 1967). But this rationalist picture of the human mind had to be drastically revised in the light of a growing number of emergent findings on irrationality: Wason's (1968) seminal studies on the inability to solve even the most simple logical reasoning problems; Goldberg's (1968, 1970) disarming demonstrations of shortcomings in expert judgments; Oskamp's (1964) early work on overconfidence; Dawes, Faust, and Meehl's (1989) provocative comparison of clinical and actuarial judgments; and most importantly of course Tversky and Kahneman's (1974) groundbreaking work on heuristics and biases. From a social psychological viewpoint, the list can be supplemented with Janis' (1972) groupthink analysis of insufficient political decision making, Weinstein's (1980) notion of unrealistic optimisms, sunk-cost (Arkes, 1985), the planning fallacy (Buehler, Griffin & Ross, 1994), and the persistence of the fundamental attribution bias (Jones & Harris, 1967).

These massive violations of rational principles are not only observed in artificial experimental reasoning tasks but in the context of highly consequential and existential problems, such as risk estimations of lethal risks, trust in expert advice, attribution of responsibility and guilt, and political decisions. While much is at stake in all these domains and irrational action is often met with enormous costs, perhaps the most scaring conclusion from countless studies is that strong motivation, incentives, and careful debriefing will often not eliminate the deficits in human reasoning.

Gullibility and the Attribution of Responsibility for Irrational Behavior

Admittedly, though, this somewhat pessimistic sketch may be not quite representative of the recent literature decision making, which is no longer concerned with costs of irrationality but with the adaptive functions of ecological and social rationality (Gigerenzer, 2000). Fast and frugal heuristics (Gigerenzer & Todd, 1999) may not accord to formal logic but nevertheless help the individual to get around in an uncertain world. Preference reversals and massive fallacies (like the conjunction fallacy) may reflect pragmatic misunderstandings of the probability concept (McKenzie & Nelson, 2003). Unrealistic optimism may be justified from the individual perspective of the beholder (Harris & Hahn, 2011). Logical reasoning seems to be intact when reasoning tasks are framed as social contracts (Cosmides & Tooby, 1992). Human perception is remarkably consistent with Bayesian calculus (Trommershausen, Maloney & Landy, 2008). And, even when no adaptive benefit is apparent, anomalies can be conceived as fully normal side effects of seemingly mild and realistic constraints of bounded rationality (Simon, 1980). No-one would contest – but nobody would take it as a dramatic deficit of the mind – that working memory is limited, that people are sometimes under-motivated, that information costs may exceed benefits from accuracy, and that optimistic biases can increase self-worth and happiness.

This perspectival shift in the literature on judgment and decision making can be characterized as a shift from irrationality to gullibility (Greenspan, 2009; Rotter, 1980; Yamagishi, Kikuchi & Kosugi, 1999) – the focus of the present volume. Gullibility is an ambivalent concept that allows for different attributions of malfunctioning and failure. Is the individual too simple-minded and naïve to solve intricate problems that exceed the individual's evolved capacities? Or does the mobilization of existing capacities depend on incentive structures and opportunity costs? Or does closer reflection reveal that apparent violations of normative rules serve some useful adaptive function? Or does failure originate in careless mistakes and

negligence of available knowledge for which the individual can be blamed? Thus, the gullibility may suggest innocence or blameworthiness, excessive demands or carelessness, external attribution to limits of bounded rationality or internal attributions to factors that are under the individual's voluntary control.

While the greatest part of the recent research emphasizes the individual's "innocence" and suggests an external attribution of mistakes and unresolved problems to the constraints of wicked environments, a different perspective is taken in the present chapter. Although there can be no doubt that illusions and biases can originate in the environment and that seemingly irrational behaviors may serve an adaptive function (Pleskac & Hertwig, 2014), this should not be misunderstood as a generalized acquittal. Serious attempts and progress in understanding the biological origins and the adaptive value of bounded rationality should not prevent us from noting clearly irrational and costly behaviors such as absurd beliefs about death panel (Nyhan, 2010), grossly biased risk estimates (Swets, Dawes & Monahan, 2000), or catastrophic losses in sunk-cost situations (Arkes, 1985). It is hard to see what sort of bounded rationality may conceal the cost and hardship caused by such misdeeds, for which a mature social agent must be blamed.

Metacognition Highlights the Individual's Responsibility

The individual is particularly responsible for those functions of quality control of one's own cognitive processes that are commonly metacognition. Metacognition is all about monitoring and control. Monitoring functions are concerned with the assessment of information validity and permissibility of cognitive operations. Control functions use the monitoring results to plan further action, basing action on valid information, discard invalid input, and correcting for biased or insufficient evidence. The remainder of this article is concerned with metacognitive myopia, a major source of rationality that originates in the individual's monitoring and control functions.

Metacognitive Myopia – Major Impediment of Rationality

A growing body of convergent evidence speaks to the noteworthy phenomenon of metacognitive myopia (MM). Pertinent research on MM (as reviewed in Fiedler, 2012) left me with the strong conviction that the ultimate reasoning deficit arises at the metacognitive level – reflecting a conspicuous failure to monitor and control for the validity of the information given, rather than at the primary cognitive level of perception, encoding, and memory functions. People are remarkably accurate in processing even complex arrays of stimulus information, and there is little evidence for restricted capacity or motivational biases as causes of strong violations of logical rules. Most striking anomalies arise in spite of sufficient capacity and perfect mastery of logical rules because people are notoriously uncritical and naïve regarding the origin and the validity of the information given. This shortsightedness (myopia) is not due to people's being insensitive to or disinterested in information but, ironically, to their being too sensitive to information, which is taken for granted though it can be suspected to be invalid and even when its invalidity is obvious.

MM at varying degrees of blatancy. Experimental evidence for MM can be organized on a continuum of varying degrees of task intricacy. On one end of this continuum are highly intricate tasks that render MM effects unsurprising, making validity problems indeed hard to understand. For instance, when observing small and large samples of behaviors (exhibited by ingroups and outgroups, respectively) hardly any layperson would take the reduced variance of smaller samples into account (Linville, Fischer & Salovey, 1989). Consequently, observers can hardly correct for an outgroup-homogeneity bias (Ostrom & Sedikides, 1992), that is, for the stereotypical tendency to perceive outgroups as more homogenous and less differentiated than ingroups.

At the other end of the continuum, the naive reliance on given evidence (hearsay, social-media tweet, gossip, advertising, others' opinions) is most conspicuous for the invalidity or

deceptive nature of stimulus input is crystal-clear. For instance, even when an explicit debriefing instruction tells people that a statement is wrong, or when people themselves correctly deny the validity of propositions and correctly recognize the input to be wrong, they nevertheless continue to be influenced by such misinformation.

Some provocative evidence to start with. Let us first illustrate the MM syndrome with some examples from social psychology, which illustrate the latter type of blatant MM effects. In a later section, we will turn to other evidence from experimental cognitive psychology, which are better suited to explain the psychological origins of the metacognitive deficit.

One striking example, to start with, can be found in Jones and Harris' (1967) seminal demonstration of the correspondence bias. Participants were asked to infer an essay writer's own political orientation from the arguments provided in an essay that was either in favor of or against the communist spirit of the Cuban leader Fidel Castro. Communist attitudes were inferred from pro-Castro essays and anti-communist attitudes from anti-Castro essays, even when participants knew that essay writers were not free to express their true opinion but were randomly assigned to either the pro- or to the anti-condition. Still, they continued to exhibit correspondent inferences (of attitudes from essay contents) even though essays were written on demand and hence fully undiagnostic.

In Jones and Harris' paradigm (1967), invalidity is obvious but it remains implicit. In a perseverance paradigm (Ross, Lepper & Hubbard, 1975), participants are explicitly debriefed of the invalidity of an alleged feedback about their test performance. Nevertheless, the perseverance effect shows that the influence of such clearly discredited fake information is not fully reversible. Participants continue to be influenced by the invalid feedback.

In two experiments reported by Fiedler, Armbruster, Nickel, Walther & Asbeck, 1996) participants who had watched a video clip of a talk show first responded to a series of questions

asking whether the protagonist had shown positive or negative behaviors (dependent on the experimental condition) vis-à-vis the other discussants. Even when they correctly denied having seen behaviors that were indeed absent from the film, their subsequent trait ratings were systematically biased towards the valence and the semantic implications of the behaviors (expressed by manifest action verbs vs. emotional state verbs) included in denied questions. Participants did not refrain from utilizing information that they had themselves correctly classified as false.

Decades later than Ross et al.'s (1975) intriguing perseverance effects, a modern research program on debunking, relying on big data from the media, conveys more or less the same message about the inability to correct for misinformation (Chan, Jones, Jamieson & Albaracín, 2017; Lewandowsky, Ecker, Seifert, Schwarz & Cook, 2012). Meta-analyses provide strong evidence for the persistence of misinformation in spite of enlightening counter messages that ought to undermine the initial belief in such myths as genetically modified mosquitoes causing the outbreak of the Zika virus in Brazil (Schipany, 2016), the existence of mass destruction weapons in Iraq before the U.S. invasion (Newport, 2013), or the alleged causal impact of the rubella vaccine on autism, measles, and mumps. The effect size of persisting misinformation in the face of debunking proved remarkable (*d* in the range of 0.75 to 1.06) in the meta-analysis provided by Chan et al. (2017).

Across all articles reviewed so far, the notion of MM was never mentioned. This conspicuous lack of interest in metacognition as a research topic in social and in applied psychology might itself be interpreted as a reflection of MM among scientists. Apparently, the need to critically monitor and control the validity of information prior to its utilization for judgments and action is hardly appreciated as a central module of adaptive behavior. In social cognition as in behavioral decision research, the individual is conceived as an agent who

processes the given stimulus data, not as a critically minded, emancipated censor who decides what information to use or to discard. Not surprisingly, Chan et al.'s (2017) meta-analysis on persisting misinformation refers to additional stimulus input as a remedy, rather than metacognitive reflection of the original input: "Persistence was stronger and the debunking effect was weaker when audiences generated reasons in support of the initial misinformation. A detailed debunking message correlated positively with the debunking effect. Surprisingly, however, a detailed debunking message also correlated positively with the misinformation-persistence effect." [p. 1531]

Understanding the Origins of MM

Maybe the aforementioned social-psychological studies are too complex and value-laden to trigger proper metacognitive reasoning. Even a pre-assigned essay may be interpreted as a reflection of the essay writer's true attitude, maybe reflecting a social norm to be authentic under constraints. Or a perseverance effect may be rationalized in terms of internally generated thoughts elicited by the initial deception. It is indeed always possible to find a plausible account for the creation of a bias in the first place, and it is hardly ever conceded that such an explanation is incomplete. In addition to the creation of a bias, a complete theoretical account must also explain why an upcoming bias is not detected by the monitoring function and why it is not corrected by the control function.

However, this failure to go beyond the mere emergence of heuristic-and-bias phenomena and to explain why initially arising biases are not detected and corrected at the metacognitive level is not only typical of social psychological research. It is also characteristic of four decades of cognitive studies on the heuristics-and-biases research program (Tversky and Kahneman's, 1974). Virtually all research is confined to anchoring, availability effects, base-rate neglect, or insensitivity to sample size (Tversky & Kahneman, 1971) as causes of initial biases; hardly any

researcher in this huge area ever tried to explain why homo sapiens does not detect and correct for the biases, or why the evolution has not equipped us with effective devices for monitoring and control.

A slightly more elaborate review of MM in cognitive psychology will provide a better understanding of origins and boundary conditions of the phenomenon. In the subsections to follow, I will particularly discuss five distinct subtypes of MM, each of which constitutes a challenging psychological research program in its own right. In the final section, I will then summarize the evidence and discuss the implications of MM research for the gullibility debate and for the attribution of responsibility in the domain of rationality.

Inability not to learn. A basic insight from a century of experimental psychology is that learning curves have positive slopes; learning strength increases with increasing number of trials. In an animal conditioning context, the more often a neutral conditional stimulus (1000 Hertz tone) is paired with the presentation of an unconditional stimulus (food), the stronger will be the conditioned reaction (saliva production elicited by the tone). In foreign language acquisition, vocabulary learning increases with repeated rehearsal. The same holds for training in sports, singing, and handcraft. That learning increases with repeated practice is not only obvious; it is inevitable. We cannot tell our mind, or our neural system, to stop learning from rehearsal. It is easy to see that we cannot tell our autonomous nervous system not to learn from repeated pairings of signals and electrical shocks. We can also not tell our memory system to stop profiting from repetition. What is experienced many times will be increasingly kept in memory. As obvious as this truism might appear, this common-sensical insight is systematically neglected in task settings that call for the exclusion of merely repeated, redundant stimuli.

Participants in a series of experiments by Unkelbach, Fiedler, and Freytag (2007) were asked to figure out how often ten different shares were among the daily winners on the stock

market. However, on some days the stock market news were presented twice on two different TV programs, so that the presentation rates diverged from the effective winning rates. For instance, given an actual wining rate of 8, the presentation rates were 8, 12 or 16. Given 6 winning days, the presentation rates were 6, 7, 9, or 12; or 4 winning outcomes were presented 4, 6, or 8 times. As participants obviously understood the task and were clearly accuracy-motivated, they could well discriminate between shares that won on eight, six, and four days, respectively. However, this did not prevent them from being misled by mere repetitions. Indeed, repetitions had a similar strong effect on the participants' evaluations of the 10 different shares as independent winning outcomes. This same finding was obtained after an explicit warning, when participants had been told that some of the stock market news would be repeated and that selective repetition could lead to misleading frequency estimates.

Given this instruction, to be sure, they might have closed their eyes during repeated news programs, or they might have monitored and assessed what shares profited from selective repetition. However, regardless of the explicit warning, they were apparently not interested in monitoring and correcting for unwarranted repetition biases. They presumably did not expect repetitions to influence their frequency estimates. It may be interesting to mention that the same biases were not just evident in numerical frequency estimates but also in ratings of the willingness to invest in the various shares.

Apparently, then, the underlying evaluative learning process is sensitive to every stimulus item presenting a share as a daily winner, regardless of whether the stimulus represents a novel winner or a repetition of an already noted winner. In any case, it is linked to a real daily winner. So, just as in Pavlovian conditioning the impact of an electrical shock is independent of whether a shock was intentional, or planned by design, the trials linking shares to winning is apparently

independent of whether the trial is novel or just a repetition. In other words, evaluative learning reflects the accrual of evaluative experience rather than a frequentist inference.

Detecting proportional changes. Support for this contention, and further evidence on why we cannot not-learn from mere repetition, comes from recent evidence on the detection of proportional change (Fiedler, Kareev, Avrahami, Beier, Kutzner & Hütter, 2016). Assessing the proportion of a focal outcome is of eminent importance in reality: Is there a change in a student's rate of correct responses, in a football team's record of successful matches, or in the acceptance rate of a political party? To investigate performance on such tasks, Fiedler et al. (2016) developed a sequential paradigm in which each trial provides participants with a binary sample of two symbols (✱ and ○). They were asked to decide whether the current sample was drawn from the same universe as the preceding sample or from different universe in which the probability p(✱) of a focal symbol ✱ had increased or decreased. Again, change judgments were sensitive to actual changes and participants were motivated to perform well. However, despite this general sensitivity to actual *p* changes, the change judgments were strongly influenced by changes in absolute sample size *n*. Increases in relative *p* were readily detected when absolute *n* increased as well, for instance, when 4✱ out of 8 increased to 10✱ out of 16. However the same proportional change from 8✱ out of 16 to 5✱ out of 8 was hardly recognized. Conversely, decreases in *p* were only noticed readily when *n* also decreased but not when absolute sample size increased. When *p* remained unchanged (i.e., successive samples were drawn from the same universe), increasing *n* misled participants to believe that *p* had increased and decreasing *n* led them to believe *p* had decreased.

These strong anomalies in the detection of change, which were replicated many times, did not reflect a failure to understand the task instructions. Participants did not simply count the cardinal number of critical ✱ symbols in the numerator. This was evident from the fact that when

n (i.e., the denominator) was held constant a change from, say, 4✱ out of 8 to 5✱ out of 8 was not different from a change from 8✱ out of 16 to 10✱ out of 16. It is also insufficient to explain the anomalies as ratio bias (Denes-Ray, Epstein & Cole, 1995) or denominator neglect (Reyna & Brainerd, 2008), because it was easy to keep *p* independent of *n* when ✱-proportions were described numerically, as normalized percentage (Fiedler et al., 2016; Experiment 2). The anomalies were only observed when proportional quantities had to be extracted inductively from an extensional sample of elementary observations.

Apparently, then, the difficulty to ignore *n* when assessing *p* arises during inductive inferences of *p* from experienced samples. And, indeed, some logical reflection shows that the inductive assessment of *p* must be sensitive to *n*. Imagine a highly motivated and perfectly unbiased participant sincerely wants to assess the proportion of ✱ symbols experienced in a binary sample, or else, the rate of pro arguments in a political debate, or different students' proportions of correct responses in the virtual classroom. In any case, the task calls for a continuous update of a sample proportion *p** of the number of focal elements *f* divided by the total number *n* of all elements. The question, however, is what positive increment or negative increment must be added to the growing sample proportion for each new observation of a focal or non-focal element. Logically, the incremental weight given to each elementary observation should be 1/*n*. In a very long list of *n* observations, each element should be given a weight of 1/100. But if *n* is only 2 or 3, the elementary observation must be given a much higher weight of ½ or 1/3, respectively.

But what should our ideally motivated, unbiased assessor do when *n* is undefined, that is, when the overall number of political votes or student answers is not known beforehand? Indeed, under most natural task conditions sample sizes *n* are unknown and uncontrollable. Nobody knows in advance how many pro and con arguments will be produced in a political debate, how

often different students in a school class raise their hands and provide responses to knowledge questions. Moreover, *n* is often fully undetermined because it is impossible to say when a sample started and when it will end. So, apart from the fact that *n* is unknown beforehand, it would be fully impossible anyway to administrate the impact of all of *n*'s belonging to hundreds of sampling tasks in which we are involved at any point in time, recalculating growing $_{n}p^*$ (after *n* observations) from the preceding estimate $_{n-1}p^*$ (after n–1 observations) according to the normative updating rule $_{n}p^* = [(1/n) \cdot n_{th}$ element$] + [(n-1)/n \cdot {}_{n-1}p^*]$. As *n* increases, such an updating rule would become more and more demanding in terms of numerical precision and replete with cumulative error.

Granting this logical dilemma, it is obviously impossible to normalize the inductive assessment of *p* for sample size *n*. Still, even though it is impossible to divide the number of focal outcomes by an exact count of *n*, our rational agent is quite sensitive to proportions, that is, he or she will somehow relate the number of focal outcomes (in the numerator) to a crude feeling (in the denominator) of *n* in the stimulus context. The same number of focal features in the current sample may thus be worth more (less) if a small (large) preceding sample suggests a smaller denominator for the current estimate.

Note that such context sensitivity might actually account for the full pattern of anomalies reported in the preceding section. A current proportion of 8✱ out of *n*=16 should be worth more if a smaller preceding sample of 4 increase from 4✱ out of n=8 suggests a smaller denominator and hence a larger increment for the encoding of the current sample elements. Thus, even when the "true or absolute *n*-change" is unknown, making a normatively perfect *p** update impossible, our notion of "context-dependent, relative *n*-changes" can account for both (a) the basic *p*-sensitivity as well as the intrusion of unwarranted *n*-effects.

Responsibility for sample-size insensitivity. The preceding causal explanation for the impossibility to control for sample size, or not to learn from extended sampling, might come like an excuse for strong anomalies in the detection of change. However, although the foregoing discussion might help to explain the inductive-learning origin of MM for sample size, it does not explain why the MM lethargy carries over to many situations in which the biasing impact and unfairness of unequal *n* is crystal-clear. Our cognitive analysis of the impossibility to assess *p* independently of *n* by no means entails an acquittal for many hard-to-belief anomalies that might be interpreted as generalized consequences of MM for unknown *n*. So let us again take a social psychological perspective on clearly irrational and unnecessary biases and shortcomings for which mature human beings must be held responsible.

One memorable example can be found in criteria based statement analysis (CBSA; Steller & Koehnken, 1989; Vrij & Mann, 2006), the diagnostic method used by expert witness to evaluate the credibility of witness reports. CBSA is typically applied in criminal (rape or sexual abuse) trials in which no physical evidence is available so that court decisions with existential consequences for the defendant depend on the validity of the credibility analysis. As the presumption of innocence implies the null hypothesis that an aggregating witness statement is wrong, the CBSA method consists in a one-sided search for linguistic truth criteria in the transcribed report. Thus, the expert witness' review and recommendation depends on how many truth criteria can be found in the report, such as amount of detail, spontaneous self-correction, or structured presentation. The CBSA count of linguistic symptoms of the veracity very often determines the diagnostic judgment and whether the defendant goes to jail for five years, loses his family and his job and his existence. However, although CBSA counts have been shown to be higher when reported experiences are real, they are also subject to a detrimental artifact, text length. A long report of 15 or 20 pages is more likely to include a reasonable number of truth

criteria than a short report of only two pages. Although so much is at stake and the flagrant bias is easy to understand and might be corrected in a straightforward manner, this problem is widely ignored in legal practice (Fiedler, 2018).

In a similar vein, the inability to perfectly monitor and control for *n* generalizes to many other situations, in which unfair judgments and evaluations depend on sample size and appropriate corrections suggest themselves: teachers' evaluations of students who provide unequal numbers of responses (Fiedler, Woellert & Tauber, 2013), self-serving biases due to larger samples of self-related than other-related experience (Moore & Healy, 2008), ingroup-serving biases, or discrimination favoring majorities over minorities showing identical rates of positive behavior (Fiedler, 2000, 2008). Or, for an example from the allegedly rational domain of science, the evidence in favor of distinct hypotheses is strongly contaminated with the biasing impact of the popularity and the number of conducted studies. Validity concerns and critical assessments of whether manipulations have been effective or whether mediation tests are logically appropriate (Fiedler, Harris & Schott, 2018) are hardly ever considered, reflecting a strong syndrome of MM in scientific practice.

Pitfalls of conditional reasoning. A similar story can be told about a long tradition of research on conditional reasoning. A cognitive analysis of the inherent difficulty suggests a rationalizing account for the conspicuous failure on such tasks, but such a rationalist account does not provide a sufficient justification for so many blatant violations of conditional-inference rules. To illustrate, consider the conditional probability $p(\text{HIV} \mid \text{positive test})$ that the HIV virus is really present among people who have been tested positively. Estimates of this conditional are highly inflated if participants are told that the baserates of HIV is $p(\text{HIV}) = 0.1\%$, that the baserate of positive test results is $p(\text{positive test}) = 1\%$, and that the reverse conditional, or hit rate of HIV-infected people who are tested positively is $p(\text{positive test} \mid \text{HIV}) = 100\%$. To provide a correct

estimate of $p(\text{HIV} \mid \text{positive test})$, Bayes' theorem prescribes that the reverse conditional must be multiplied with the ratio of the two baserates:

$p(\text{HIV} \mid \text{positive test}) = p(\text{positive test} \mid \text{HIV}) \cdot p(\text{HIV})/p(\text{positive test}) = 100\% \cdot 0.1\% / 1\% = 10\%.$

That most judges grossly overestimate this surprisingly remarkably low figure is easy to "explain" or to justify, by simply admitting that lay people are not in full command of Bayesian calculus. Understanding that the ratio of two inverse conditional probabilities $p(\text{HIV} \mid \text{positive test}) / p(\text{positive test} \mid \text{HIV})$ is identical to the ratio $p(\text{HIV})/p(\text{positive test})$ of corresponding base rates sounds like higher mathematics that ordinary people cannot be expected to be understood. So nobody would come to blame ordinary people who dramatically overestimate risks expressed as conditional probabilities.

However, again, the strong anomalies persists under conditions that make it easy and obvious to overcome the underlying base-rate neglect (Bar-Hillel, 1984), or the failure to consider the ratio of baserates. For instance, consider an experiment (cf. Fiedler, Brinkmann, Betsch & Wild, 2000) in which participants who know the HIV baserate is very low (say, 1 out of 1000) are presented with an index-card file with two slots, one containing very few HIV cases and another slot with a huge number of (1000 times more) not-HIV cases. Each index card has the diagnosis (HIV vs. not HIV) on one side and the test result (positive vs. negative) on the other side. Participants can sample as many cards from the file as they feel appropriate to make an accurate estimate of $p(\text{HIV} \mid \text{positive test})$. A typical search pattern would be that participants sample all (rare) HIV cases plus a similar number of not-HIV cases. Noting that the test result is positive for 100% of all HIV cases but for hardly any sampled not-HIV cases, and from this comparison infer that a positive test result is a very good predictor of HIV.

The serious flaw does not lie in a failure to apply Bayesian calculus. It lies in the obvious fact that their sample of roughly equal numbers of HIV and non-HIV cases is extremely biased.

While it contains all HIV cases, it only contains a vanishingly small subset of not-HIV cases. Basing a conditional estimate of HIV on such a dramatically biased sample, which over-represents the true rate of HIV cases by many thousand percent, is reflective of an incredibly blatant version of MM, for which mature adult people can be held responsible. Independent of intelligence and age, it is easy to understand that a raffle that selectively includes all unattractive items is not attractive or that a forecast of a football team's winning record should not rely on a sample that draws heavily on only the worst matches in the previous season.

Yet, as a consequence of MM, people continue to make important inferences from such extremely biased samples, the invalidity of which is not at all beyond the scope of human intelligence. Even when given a forced choice between a biased and an unbiased sample, MM often prevents homo sapience from choosing correctly. Thus, keeping with the previous example, when intelligent participants (e.g., highly educated students) can choose between (1) a biased sample that contains equal numbers of HIV and not-HIV cases and (2) an unbiased sample that contains proportionally more not-HIV cases, they typically prefer to base their estimates on the former sample. Apparently, MM gives more attention to the superficial convenience of an equal-*n* design than to a critical or thoughtful check on whether the very attribute to be estimated is misrepresented in the sample.

Impoverished causal reasoning. Causal impact judgments provide another example of MM that at first sight appears to reflect an adaptive property of the human mind. Logically, the impact of a manipulated change $\triangle x$ in a causal condition x on an observed change $\triangle y$ in an effect dimension y can be quantified by a ratio $\triangle y / \triangle x$. Thus, causal impact is highest when a maximal effect (high numerator) is brought about by a minimal causal input (small denominator). For instance, if a very small dosage of a poison is sufficient to kill a huge elephant (strong $\triangle y$), the

causal impact is higher than if the same dosage only kills a tiny mouse, or if a much higher dosage (larger $\triangle x$) is required to kill an elephant.

This ratio principle underlying the notion of causal impact – dividing the size of an effect by the amount of causal input that was required to produce the effect – appears plausible and not too complicated for the human mind. It is also evident that the principle is rational and functional for the solution of many practical problems. Given that 10 grams of a substance have a nutrition value of 50 calories, we can infer than 100 grams of the same substance have 500 calories. Or, if 5 grams of another substance have 50 calories, its nutrition value must be twice as high. Yet, actual judgments of causal influences are not sensitive to this uncontested ratio principle. Real judgments are often exclusively sensitive to effect sizes and largely ignore the amount of necessary causal input. MM leads us to take the experienced sample for granted and not to think much about the causal story behind. We assess whether a patient's depression is mild or severe, but hardly ever evaluate the degree of depression relative to the extremity of stressors and loss experiences in that patient's life (Krantz, 1988). When we note how much Manchester City dominates the English Premier League in soccer, we do not relate their achievement to the incredibly high financial input that was necessary to produce the success. Or, in science, we praise studies with high effect sizes, but we hardly ever consider how strong an experimental manipulation was necessary to produce an effect. And, we definitely do not downgrade an effect ($\triangle y$) if the causal input or treatment strength ($\triangle x$) was too high. It is hardly possible to publish a study when effects are too weak, but reviewers and editors will not reject a paper if the causal treatment was too strong.

Thus, whenever causal origins cannot be fully ignored as in science, researchers do not seem to apply the compelling ratio principle but they base their evaluations on the covariance principle: An experiment is considered ideal if a strong cause produces a strong effect, not if

weak input managed to produce strong input. The same holds, by the way, for lay judgments of causal impact (Hansen, Rim & Fiedler, 2013): If “45-minute waiting time causes an increase in customer anger of 10 scale points”, the subjective causal influence is stronger than if “14-minute waiting time causes an increase in customer anger of 10 scale points”.

Again, one might explain and justify the covariance principle as adaptive and more influential than the ratio rule. Such an explanation (Fiedler et al., 2012) might start from the assumption that in reality adaptive agents not only face the task of quantifying the impact of a single cause, but they also have to detect the influence of a cause in the context of many different causes *a*, *b*, *c*, … *w* that vary at the same time. When in such a multi-causal setting an effect $\triangle y$ co-occurs with changes in several causal factors, a very subtle change in, say, $\triangle a$, will be much less detectable than a massive change in, say, $\triangle w$. A loud and attention-grabbing provocation is more detectable and will thus appear to have a stronger causal impact on an aggressive act or crime than a hardly detectable, subtle insulting gesture.

While covariance looks like an adaptive rule in the context of detectability, it hardly justifies the maladaptive neglect of the ratio principle and the wide-spread tendency to focus on salient effects while ignoring causal origins. Just as a developmentally higher levels of moral judgments are sensitive to underlying intentions and voluntary control and not only to the amount of damage caused by moral transgressions (cf. Kohlberg, 1994), rational judgments of causality must relate the strength of effects to the strength of causal input.

Divergent trends at different aggregation levels. Last but not least, a prominent final example is MM for existing differences between aggregation levels. On one hand, the vicissitudes of aggregation levels are intrinsically counter-intuitive, and one is tempted to excuse their neglect. It is hard to understand, for instance, that the correlation between Black skin color and illiteracy is negligible at the level of individual people but close to perfect at the level of

large geographic districts. Black individuals are hardly more likely than White people to be illiterate, but the correlation between the proportion of Black people and the proportion of illiterates in different US districts is very high ($r > .80$ for very large districts). This is difficult to understand; the genetic influences underlying individuating correlations are fully independent of the economic factors producing the ecological correlation at high district level (Robinson, 1980).

However, on the other hand, the notion of divergent trends at varying aggregation levels does not exceed our intelligence and we are familiar with many pertinent examples. Rich nations may have high poverty rates; what is pleasant in the short run may be unpleasant in the long run; or research findings obtained at group level need not hold for individual participants. Despite the counter-intuitive, intricate nature of aggregation level effects, the human mind is not bound to myopia or blindness for aggregation levels. As nicely illustrated by Eagly and Steffens' (1984), the gender stereotype that leadership ability is typical of males but not of females is true at the high aggregation level of vocational environments. In vocational fields with the highest rate of leadership ability (top management in organizations) the rate of male people is highest. It is nevertheless possible that, at the level of individuals, the few female leaders working in the top management outperform the majority of male leaders, thus creating a zero correlation or an inverse correlation at individuating level.

Conclusions: Gullibility, Myopia, and Social Responsibility

Thus, a review of MM effects in various paradigms reveals that rationality research is intimately related to the ambivalent concept of gullibility, the meaning of which implies both innocence and negligence. A good deal of recent work on judgment and decision making highlights the normal origins and the adaptive functions served by many apparent violations of rational norms. Biases and illusions can be explained as normal consequences of ordinary laws of learning, properties of the probabilistic environment, and intrinsic difficulties of some inference

tasks. However, even when the origins of irrational behavior can be understood and rationalized, this does not exempt the individual of his or her social responsibility. In spite of bounded rationality, the individual remains blameworthy. Not all deficits in metacognitive functions can be attributed to unavoidable constraints. Some MM effects are unbelievably blatant and naïve and not enforced by task demands that exceed our cognitive capacities. We perfectly understand that selective repetition may bias impressions, that the probability that males are millionaires is much lower than the probability that millionaires are male, that highly concentrated poison has more causal power than diluted poison, or that happiness of nations is not happiness of people. And yet, we fall prey to repetition biases, fail on highly meaningful conditional reasoning tasks, we misunderstand the ratio principle of causal inference, and we are completely confused by divergent trends observed at different aggregation levels.

The MM perspective on rationality points to missed opportunities to utilize insights and critical analyses that are easily understood and hard to contest. It appears as if, for some reason, we are simply not interested or motivated to engage in critical assessment, or to cast the validity of a flawed sample into question. For an illustrative example, consider the recent “me too” debate on the social media, and its echo in the mass media. Regardless of what part of the information solicited in this public debate is true, semi-true, exaggerated, or even faked, and regardless of whether part of the reported transgressions are harmless and manifestations of normal mating behavior, the sampling procedure or “research design” underlying this media game is sorely biased. The debate relies on a retrieval prompt that exclusively refers to the worst exemplars of norm-violating behaviors represented in the extreme part of the distribution of (male) human conduct. The amount and strength of evidence solicited by such a sampling process does not tell us anything about the relative rate of such misbehaviors, because normal and nice behavior is ignored. All we can infer from such a lop-sided sampling process is that a large number of social-

media agents (maybe more than a billion) have been reached by the “me-too” prompt. Although we fully understand that such a “research design” is useless, we nevertheless continue to be impressed by the pessimistic results. It is somehow comparable to persisting optimal illusion that continues to fool our perception in spite of perfect debriefing.

Nevertheless, an ultimate goal of a gullibility debate must be to counter this MM lethargy and to remind people of their responsibility to engage in critical monitoring and control. Neither restricted working memory nor lack of incentives nor any other aspect of bounded rationality restricts of ability, and our obligation to monitor and control the quality of the information that impinges on our mind. For some inference problems there may be no patent remedies at the metacognitive level. Repetition biases cannot be turned off, the baserates needed to deal with conditional inference problems may be unknown, and information may not be available at the appropriate aggregation level. Even then, however, we can still recognize dangerous situations in which a stimulus sample is flawed, an information source is untrustworthy, or unequal sample size must lead to unfair and lop-sided comparisons. And we can decide not to act or to discard information that is obviously flawed.

Recent work on nudging (Thaler & Sunstein, 2008) and prudent default setting (Johnson & Goldstein, 2003) emphasizes environmental design and external decision aids as key interventions. The MM perspective suggests an opposite, self-determined and internally controlled approach, namely, critical assessment and emancipation at the metacognitive level. Which of the two opposite approaches turns out to be superior is a matter of future research, but for the moment, the gullibility debate can help to articulate the psychological underpinnings of both positions.

References

Arkes, H. R., & Blumer, C. (1985). The psychology of sunk cost. *Organizational Behavior And Human Decision Processes*, *35*(1), 124-140.

Bar-Hillel, M. (1984). Representativeness and fallacies of probability judgment. *Acta Psychologica*, *55*(2), 91-107.

Buehler, R., Griffin, D., & Ross, M. (1994). Exploring the 'planning fallacy': Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, *67*(3), 366-381.

Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, *28*(11), 1531-1546.

Cosmides, L. (1989). The logic of social exchange: has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31, 187-276.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptions for social exchange. In J.H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adaptive mind: Evolutionary psychology and the generation of culture* (pp. 163-228). New York: Oxford University Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243*, 1668–1674.

Denes-Raj, V., Epstein, S., & Cole, J. (1995). The generality of the ratio-bias phenomenon. *Personality And Social Psychology Bulletin*, *21*(10), 1083-1092.

Denrell, J., & Le Mens, G. (2012). Social judgments from adaptive samples. In J. I. Krueger (Eds.) , *Social judgment and decision making* (pp. 151-169). New York, NY, US: Psychology Press.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women

and men into social roles. *Journal of Personality and Social Psychology, 46*(4), 735-754.

Fiedler, K. (2000a). Beware of samples! A cognitive-ecological sampling theory of judgment biases. *Psychological Review, 107*, 659-676.

Fiedler, K. (2000b). On mere considering: The subjective experience of truth. In H. Bless, J. P. Forgas, H. Bless, J. P. Forgas (Eds.) , *The message within: The role of subjective experience in social cognition and behavior* (pp. 13-36). New York, NY US: Psychology Press.

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

Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical inference. *The Psychology of Learning and Motivation, 57*, 1-55.

Fiedler, K. (2018). In R.S. Sternberg (Ed.), A missed opportunity to improve on credibility analysis in criminal law. *My biggest research mistake*. Sage.

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*, 399-418.

Fiedler, K., Kareev, Y., Avrahami, J., Beier, S., Kutzner, F., & Hütter, M. (2016). Anomalies in the detection of change: When changes in sample size are mistaken for changes in proportions. *Memory & Cognition, 44*(1), 143-161.

Fiedler, K., Armbruster, T., Nickel, S., Walther, E., & Asbeck, J. (1996). Constructive biases in social judgment: Experiments on the self-verification of question contents. *Journal of Personality and Social Psychology, 71*(5), 861-873.

Freytag, P., & Fiedler, K. (2006). Subjective Validity Judgments as an Index of Sensitivity to Sampling Bias. In K. Fiedler, P. Juslin, K. Fiedler, P. Juslin (Eds.) , *Information sampling*

and adaptive cognition (pp. 127-146). New York, NY US: Cambridge University Press.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*. New York: Oxford University Press.

Gigerenzer, G., & Todd, P. M. (1999). Fast and frugal heuristics: The adaptive toolbox. In , *Simple heuristics that make us smart* (pp. 3-34). New York, NY, US: Oxford University Press.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious*. New York: Viking Press.

Goldberg, L. R. (1968). Simple models or simple processes? Some research on clinical judgments. *American Psychologist, 23*(7), 483-496

Goldberg, L. R. (1970). Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin, 73*(6), 422-432.

Greenspan, S. (2009). *Annals of gullibility: why we get duped and how to avoid it*. Praeger Publishers.

Harris, A. L., & Hahn, U. (2011). Unrealistic optimism about future life events: A cautionary note. *Psychological Review, 118*(1), 135-154.

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135*, 555–588.

Janis, I. L. (1972). *Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes*. Oxford, England: Houghton Mifflin.

Jones, E. E., & Harris, V. A. (1967). The Attribution of Attitudes. *Journal of Experimental Social Psychology, 3*(1), 1-24.

Jones, E. E., & McGillis, D. (1976). Correspondent inferences and the attribution cube: A comparative reappraisal. *New directions in attribution research, 1*, 389-420.

Kelley, H. H. (1967). Attribution theory in social psychology. *Nebraska Symposium on Motivation, 15,* 192-238.

Kohlberg, L. (1994). The claim to moral adequacy of a highest stage of moral judgment. In B. Puka, B. Puka (Eds.) , *The great justice debate: Kohlberg criticism* (pp. 2-18). New York, NY, US: Garland Publishing.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest, 13*, 106–131.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology, 57*(2), 165-188.

McKenzie, C. M., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review*, *10*(3), 596-602.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502-517.

Newport, F. (2013). *Americans still think Iraq had weapons of mass destruction before war*. Retrieved from http://www.gallup.com/poll/8623/americans-still-think-iraqhad-weapons-mass-destruction-before-war.aspx

Newport, F. (2015). *In U.S., percentage saying vaccines are vital dips slightly*. Retrieved from http://www.gallup.com/poll/181844/percentage-saying-vaccines-vital-dips-slightly.aspx

Nyhan, B. (2010). Why the “death panel” myth wouldn’t die: Misinformation in the health care reform debate. *The Forum*, *8*(1), Article 5.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal Of Consulting Psychology*, *29*(3), 261-265.

Ostrom, T. M., & Sedikides, C. (1992). Out-group homogeneity effects in natural and minimal groups. *Psychological Bulletin*, *112*(3), 536-552.

Piaget, J. (1950). *The psychology of intelligence*. Oxford, England: Harcourt, Brace.

Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, *143*(5), 2000-2019.

Johnson, E. J., & Goldstein, D. G. (2003), "Do Defaults Save Lives?" *Science*, *302* (5649), 1338-39.

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, *18*(1), 89-107

Ross, L., Lepper, M.R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attribution processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32*, 880-892.

Rotter, J.B. (1980). Interpersonal trust, trustworthiness, and gullibility. *American Psychologist*, *25* (1), 1–7.

Sarbin, T. R., Taft, R., & Bailey, D. E. (1960). *Clinical inference and cognitive theory*. Oxford, England: Holt, Rinehart & Winston.

Schipani, V. (2016). *GMOs didn't cause Zika outbreak*. Retrieved from http://www.factcheck.org/2016/02/gmosdidnt-cause-zika-outbreak/

Simon, H.A. (1982). *Models of bounded rationality*. Cambridge, MA: MIT Press.

Sperber, D. (2009). Culturally transmitted misbeliefs. *Behavioral & Brain Sciences, 32*, 534–535.

Steller, M., & Köhnken, G. (1989). Criteria-based statement analysis. Credibility assessment of children's statements in sexual abuse cases. In D.C. Raskin (Ed.), *Psychological methods for investigation and evidence* (pp. 217-245). New York: Springer.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest*, *1*(1), 1-26.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. New Haven, CT, US: Yale University Press.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision making, movement planning and statistical decision theory. *Trends in Cognitive Sciences, 12*, 291–297.

Tversky, A. & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105-110.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Vrij, A., & Mann, S. (2006). Criteria-Based Content Analysis: An empirical test of its underlying processes. *Psychology, Crime & Law, 12*(4), 337-349.

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*, 37-52.

Wason, P. C. (1968). Reasoning about a rule. *The Quarterly Journal of Experimental Psychology, 20*(3), 273-281.

Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal Of Personality And Social Psychology, 39*(5), 806-820.

Yamagishi, T., Kikuchi, M. & Kosugi, M. (1999). Trust, gullibility, and social intelligence. *Asian Journal of Social Psychology, 2* (1), 145–161.