THE BEST SWIMMERS DROWN – MECHANISMS AND EPISTEMIC RISKS: A CONSTRUCTIVE CRITIQUE OF ELSTER

Johannes Persson

Department of Philosophy, Lund University

Johannes.Persson@fil.lu.se

ABSTRACT: According to Jon Elster, mechanisms are frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences. In the absence of laws, moreover, mechanisms provide explanations. In this paper I argue that Elster's view has difficulties with progressing knowledge. Normally, filling in the causal picture without revising it should not threaten one's explanation. But this seems to be Elster's case. The critique is constructive in the sense that it is built up from a discussion of a mechanism that might explain 'unwarranted' risk taking in connection with swimming—a mechanism that is mirrored in the proverb: The best swimmers drown.

-N'oubliez pas, je suis meilleure que vous à ce jeu-là.

-Peut-être, mais ce sont toujours les meilleurs nageurs qui se noient.

1. Introduction

This paper has a twofold aim. On the one hand it is a constructive critique of Jon Elster's account of mechanisms in general and of type B-mechanisms, i.e. mechanisms with indeterminate consequences, in particular. On the other hand I promote the idea of epistemic risks, i.e. risks primarily concerning brittle knowledge. The naturalness of this manoeuvre has perhaps never been more transparent than this unusually warm summer when Swedish newspapers have been flooded with reports of drowning accidents. One of Elster's illustrations of type-B mechanisms involves the explanation of drowning accidents, and one of the reasons his account of mechanisms is problematic is easy to locate when we in more depth consider reported causes of drowning. A special kind of risk taking, *epistemic risk taking*, seems to be involved in many drowning accidents, and this furthering of our knowledge creates unexpected difficulties for Elster's account. There is one more reason why the example this text revolves around deserves attention. The explanation of drowning accidents is indeed an important risk question in its own right.

This topic, I trust, will connect eventually to Wlodek Rabinowicz' interest in causal decision making and evidentiary value mechanisms. I am not there yet. In the meantime I hope to continue learning from Wlodek's always helpful comments.

2. "The best swimmers drown" and the laws

In Christopher Hampton's play *Les liaison dangereuses* the best swimmers drown.[1] A similar expression is often reported as a proverb of Italian origin. I have also seen it listed in an African compilation of proverbs, according to which it has its source in Luyia,Western Kenya. The exact phrasing of the proverb differs, though, and when *Les Liaison dangereuses* was recently staged by The Shakespeare Theater of New Jersey it was formulated in quite different ways by the reviewers. Compare, for instance:

> While the Vicomte admits, "It's *only* the best swimmers who drown," the innocent victims who are to be submerged in the lugubrious cesspool are legion.[2]

> Yet hearts do break, and *even* the best swimmers drown. […] 'Les Liaisons Dangereuses' is at the Shakespeare Theater of New Jersey, 36 Madison Avenue, Madison, through July 24.[3]

One of the things Jon Elster is well known for is that he takes the explanatory power of proverbs seriously. On the surface, proverbs may seem to express rather naïve lawlike regularities and conform to the philosophically most well-known type of explanation, the covering law:

---

[1] The play is based on Chaderlos de Laclose's classic novel, but I haven't discovered the proverb in Laclose.

[2] Review in *Theatre scene*.net, my italics.

[3] Review in *The New York Times*, my italics.

(1) All the best swimmers drown

(which together with the fact that someone was a good swimmer explains why he drowned). But, as Elster points out, even if we put this relationship in terms of increased probabilities and risks rather than in terms of necessitation of actual manifestations,[4] this is not a fruitful way to develop the analysis:

> The proverb "The best swimmers drown" would be absurd if taken to mean that the propensity to drown invariably increases with swimming skill.[5]

Equally absurd, we may add, would be the statement expressed as a necessary condition, as one of the above reviewers have it:

(2) Only the best swimmers drown

(1) and (2) are obviously not true. Like many other popular and scientific expressions, proverbs cannot be understood as the attempted covering law explanations in (1) and (2). Not only would they be deemed false (or never actual, only potential[6]) by such an account, they would come out much too

---

[4] Plenty of interesting problems emerge with such a probabilistic move, but these problems are not the focus of this paper.

[5] Elster (forthcoming 2007)

[6] To use Hempel's sometimes valuable distinction.

immature too. These expressions need to be interpreted in ways that do them better justice. But doing proverbs justice is not only a matter of keeping them true. It is equally important that their explanatory power is preserved. Hence, it is not good enough to go from 1) or 2) to:

(3) Some of the best swimmers drown

Although true, the explanatory power of (3) is almost zero. (3) seems rather to point to the need for an explanation than function as one itself. This seems to be the case even in the most fortunate situation, where a contrastive situation such as the following is assumed: A, the good swimmer, drowns whereas B, the moderately skilled swimmer, survives. "Why did he drown? He was such a good swimmer"; "Well, some of the best swimmers drown". (1) and (2) would have succeeded better as explanations in this case, even if both probably would have been followed by a new explanatory-seeking question: "How is it possible that good swimmers drown?". (3) rather motivates the repeated explanatory-seeking question: "But why did he drown?". (3) results in no explanatory progress at all.

3. Varieties of mechanisms and mechanistic explanation

Since (1)-(3) are inappropriate, the assumed explanatory power of proverbs implies that there is a distinct kind of explanation in addition to covering law explanations. For Elster, covering laws is still the ideal, and it is obvious that mere descriptions won't do. This is why he thinks of the kind of explanation needed as occurring on a level between covering law

explanation and mere descriptions of things. Elster returns to this explanatory middle level on many occasions in his later books and articles on explanation. He is by no means only interested in the explanatory power of proverbs. They are used as illustrations of a kind of conclusion that can be found in many of the special sciences where laws are rare:

> Are there lawlike generalizations in the social sciences? If not, are we thrown back on mere description and narrative? In my opinion, the answer to both questions is no. The main task of this chapter is to explain and illustrate the idea of a *mechanism* as intermediate between laws and descriptions.[7]

According to Elster, the middle level can be inhabited by mechanisms and be governed by mechanistic explanations. But what is a mechanism, and what constitutes a mechanistic explanation? The literature on mechanisms scouts a multitude of paths. Surprisingly few have been fully charted, and efforts of integration are rare. Thus, so far "mechanistic explanation" can mean a lot of things, depending on whether, for instance, *production* (mechanisms produce Humean laws and their instances), *complex systems* (mechanisms are structured and it is in virtue of this that they produce and explain), or *the hidden tie* (mechanisms supply the missing link between cause and effect) has been the focus of mechanistic interest. Wesley Salmon belonged to the first camp. For Salmon, causal processes are the "mechanisms that propagate structure and transmit causal influence in this

---

[7] Elster (1999), p. 1.

dynamic and changing world."[8] What is important about mechanisms, according to this conception, is their power to produce. Stuart Glennan focuses instead on complexity. According to him, a mechanism consists of a number of those parts. It operates by the interaction of parts. A mechanism for a behaviour is a complex system that produces that behaviour by the interaction of a number of parts.[9] Peter Railton, Lindley Darden, and John L. Mackie can be seen to exemplify the third direction as all three are concentrating on mechanisms as the hidden tie between cause and effect. Railton says that an account of a mechanism is "a more or less complete filling-in of the links in the causal chains."[10] The following quote from Darden is also representative for this direction: "Complete descriptions of mechanisms exhibit productive continuity without gaps from the set up to the termination conditions, that is, each stage gives rise to the next."[11] Mackie, finally, said that a causal mechanism is "some continuous process connecting the antecedent in an observed […] regularity with the consequent."[12]

I do not want to imply that these three directions are exhaustive of mechanistic theorising—in a moment we will see that Elster follows a path of his own—nor are the three paths impossible to pursue simultaneously. The quote from Darden, for instance, nicely illustrates the production idea

---

[8] Salmon (1997), p. 66.

[9] Glennan (2002), p. 344.

[10] Railton (1978), p. 478.

[11] Darden (2002), p. 356.

[12] Mackie (1974), p. 82.

as well. I have no intention to penetrate all these varieties and combinations of mechanisms and their corresponding notions of mechanistic explanation. For the moment, the above display simply serves as proof that the diversity of mechanistic conceptions requires some specification. What kind of mechanistic explanation do we have in mind when we locate mechanisms between laws and plain facts? Since I will pursue the Elster-trail here his approach will provide the guidelines. But the choice is motivated by less private reasons as well. Elster's influence especially on the social sciences should not be underrated. It is an important account and use of mechanisms.[13] Here is a statement of Elster's present view of mechanisms:

> Roughly speaking, mechanisms are frequently occurring and easily recognizable causal patterns that are triggered under generally unknown conditions or with indeterminate consequences.[14]

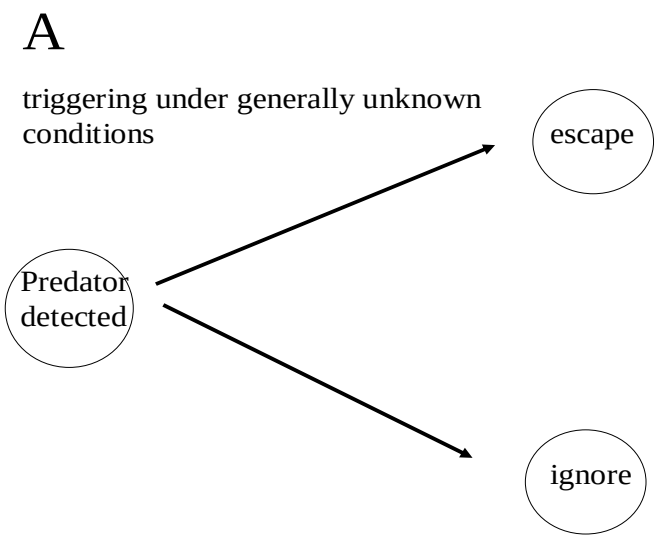
The first part about the frequently occurring and easily recognizable guarantees the kinship with traditional covering laws. The second part distinguishes mechanisms from such laws. Moreover, the second part divides naturally into two kinds of situation, type A and type B.

---

[13] Elster's understanding of mechanisms may have changed too. At least his focus has. From having been interested in the effects of making explanations finer grained, his later work is more concerned with explanations in the absence of covering laws. Here I am concerned with his later work, such as it appears at least from "A plea for mechanisms" in Hedström and Swedberg (1998) and onwards.

[14] Elster (forthcoming 2007).

In a situation of type A, either this or that process is triggered. The prey detects the predator. Sometimes this leads to escape, but sometimes the prey ignores the predator. Type A has to do with what happens in the start, it is about causes.

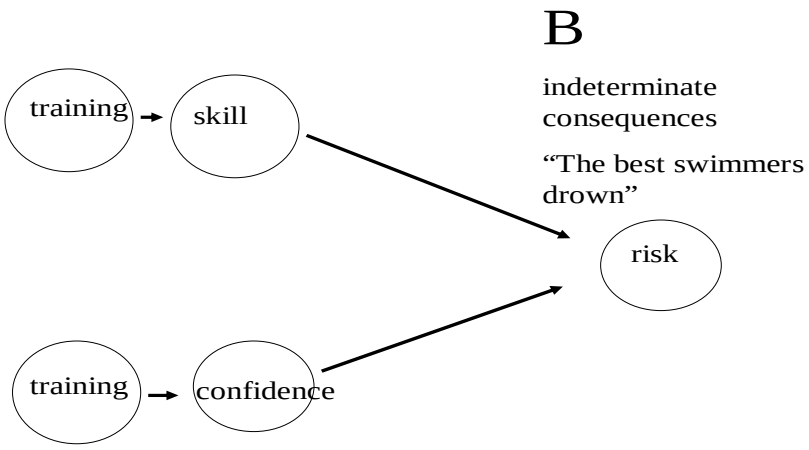


In a situation of type B, two processes are triggered with indeterminate consequences. Type B has to with what happens in the end, it is about effects. It is such a situation Elster thinks the proverb is intended to capture. It is only that one of the causal processes is taken for granted and not reported:

For all swimmers, probably, training increases swimming skill and increases confidence in swimming skill too. Both affect the person's risk taking. For some swimmers confidence in their swimming skill increases

more rapidly than their skill, causing them to take unwarranted risks. This constitutes Elster's attempt of accounting for the explanatory power of the proverb we are interested in. This is compatible with the fact that for some swimmers it is the other way round, and also with the possibility that some swimmers are more or less perfectly calibrated:

> The proverb 'The best swimmers drown' would be absurd if taken to mean that the propensity to drown invariably increases with swimming skill. Yet for some swimmers it may indeed be the case that their confidence in their swimming skill increases more rapidly than their skill, causing them to take unwarranted risks ('Pride goes before a fall').[15]



[15] Elster (forthcoming 2007).

It is clear that the mechanistic picture we are presented with in figure B is more explanatory than an expression like “some of the best swimmers drown” (i.e. 3). By representing the causal processes involved, as in B, we achieve much better understanding of the causal pattern than if we, as in 1)-3), approach the situation from a pure perspective of law or regularity. Finally, and for the sake of the argument, I will grant that we would have understood even more had we had proper, exceptionless laws to rely on. Elster’s mechanistic explanations on their level between covering laws and mere descriptions seem indeed to be in the right place.

4. The many drowning accidents in Sweden

I reside among those, may it be David Hume or the Swedish poet Erik Beckman, who think that we should primarily concern ourselves with that which is factive. Philosophy dealing only in possibilities is rather unattractive. A type B situation, such as the best swimmers drown, can easily be constructed but deserves to be taken seriously only if it is a viable explanation.

It is therefore of interest to see what the experts say. Swedish Life Saving Society presents monthly reports of drowning accidents in Sweden. For instance, in July this year 46 people drowned in Sweden. Poor ability and overconfidence were both cited as possible explanations in subsequent analyses. In fact the relationship between earlier training and overconfidence was frequently pointed out as a cause:

> Many probably overestimate their ability. Often it is several years since they acquired simborgarmärket [a proof that one has swum 200 metres in deep

> water]. When bathing they usually swim out and in again. They swim 25 metres maximum.[16]

Middle aged and older men in particular seem to be vulnerable in this respect:

> It is beliefs like 'happens to others but not me', says Anders Wernesten. He claims that many, men especially, are overconfident about what they manage in natural waters. A person who has swum 2000 metres once and in a well tempered pool believes he has the same capacity in colder and unsafe lakes and seas.[17]

"Män simmar rakt ut—kvinnor simmar längs stranden" ("Men swim into open water—women swim along the shore"), says one of the articles in the electronic version of *Dagens Nyheter*, August 24 2006. Similarly, drowning accidents during boat fishing in small lakes happen quite often.[18] And apparently not only in Sweden:

> Even the best swimmer can drown if they're knocked out by a swinging boom on a sailboat, or if they lose their grip on an overturned boat because the water is too cold.[19]

---

[16] My translation. *DN.se* July 9 2006.

[17] My translation. *Svenska Dagbladet* July 31 2006

[18] *Drunkningsolyckor 2005*, p. 3.

[19] *CBS News Online* July 19 2005

> Even the best swimmers can misjudge the water and their skills when boating or fishing; conditions change quickly in open water.[20]

Hence Elster's mechanistic explanation seems to be on the right track, and this makes this specific example all the more interesting to analyse further.[21] Where then does the supposed difficulty for his analysis emerge? My following criticism builds on what happens when our knowledge is furthered in a particular way which I will establish first (section 5-7). The immediate consequence of this aggregation of knowledge will be that we distinguish two effects, two kinds of risk, in Elster's mechanistic explanation. This progress will create unexpected difficulties to be examined closer in a while (section 8).

5. Outcome risk and epistemic risk

In the following I will adopt a distinction between two kinds of risk promoted by Nils-Eric Sahlin and his colleagues[22]:

Understanding of risk can be, and frequently is, directed immediately towards consequences or outcomes of decisions or actions, or—as Sahlin

---

[20] European Child Safety Alliance, p. 2

[21] Since Swedish Life Saving Society complains that a good explanation of the fact that there is a difference in behaviour between men and women is lacking, some of us might find such a task especially challenging. See for instance the interview with Lothar Schelp in ”män och olyckor hänger ihop”, DN.se, August 23 2006.

[22] Gärdenfors and Sahlin (1982/1988), Sahlin and Persson (1994).

and Persson (1994) has it—consequences of certain well-defined events. What is at issue then is "outcome risk". Theory of risk has often been equated with theory of outcome risk.

But the idea of outcome risk does not capture all risks. Understanding of risk can be, and frequently is, directed primarily towards the decision maker's or agent's knowledge of the consequences or outcomes of decisions, actions, or well-defined events. What is at issue then is "epistemic risk".

Imagine someone considering playing the roulette—often a risky business indeed. In light of the distinction, we can identify two varieties of risk taking roulette players: those who assume that the set up is fair and those who suspect that it might have been tampered with in some way. The first variety would identify themselves as outcome risk takers only, but the second variety consider epistemic risk taking too.

I will not enter a discussion of the exact relationship between epistemic risks and outcome risks here, but it is clear that in cases where epistemic risks are considerable, outcome risks become difficult to even monitor.[23] Hence despite the academic preoccupation with outcome risks there is a sense in which epistemic risks are more fundamental than outcome risks.

Does the distinction apply in the present context? Let us compare the two warnings from CBS and European Child Safety Alliance:

---

[23] Compare Sahlin and Persson (1994), p. 38.

(4) Even the best swimmer can drown if they're knocked out by a swinging boom on a sailboat, or if they lose their grip on an overturned boat because the water is too cold.

(5) Even the best swimmers can misjudge the water and their skills when boating or fishing; conditions change quickly in open water.

(4) is clearly about the consequences of certain well-defined events, the swinging of a boom and the overturn of a boat. The explanation doesn't involve anything concerning the state of knowledge or reflection of the exposed. (4) is about outcome risk only. (5) might be different. It has a partial content at least that involves the decision maker or agent and his or her reflection on the situation. A full fledged analysis would show that (5) involves epistemic risk also. Hence the distinction applies to the present context. Both epistemic and outcome risks figure in explanations of the best swimmers' drowning accidents. Next I will show how the two kinds of risk enter Elster's analysis.

The first observation is plain enough. Elster's first causal process involves ability. Training increases ability and, whether or not the swimmer reflects on this matter, ability decreases risk. This causal process is on level with (4). It seems evident that this causal process leads to decreased *outcome* risk.

6. Confidence and epistemic risk

Elster's second causal process involves confidence. The explanation as a whole builds on the idea that confidence is sometimes poorly calibrated.

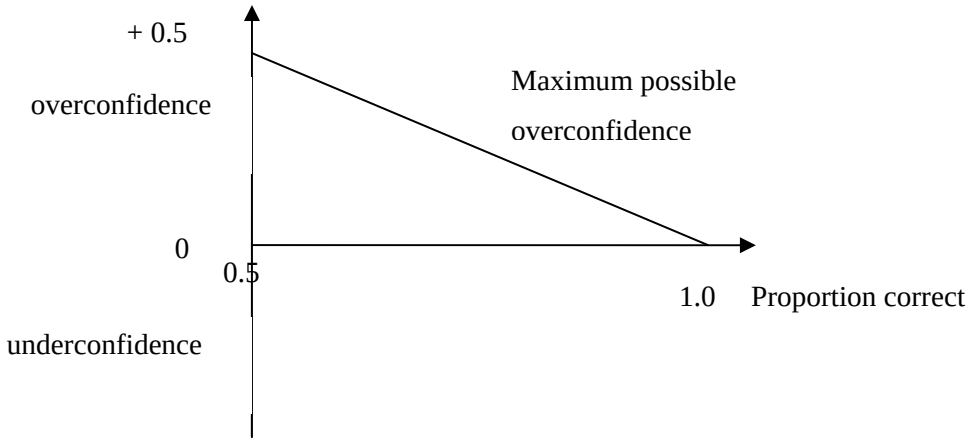
The fact that the best swimmers drown, according to Elster, has to do with *overconfidence*—for some swimmers confidence in their swimming skill increases more rapidly than their skill.

What is overconfidence? In psychology over (and under) confidence is frequently measured in a relatively straightforward manner, which I think is satisfactory enough for our purposes. Compute the mean subjective probabilities (x) assigned to the correctness of answers minus the proportion of correct answers (c). In other words confidence measures how certain you are that you are correct in your opinion. If we assume that subjective probabilitites have something to do with how we settle for an opinion in the first place, confidence is a second-order measure. A person is overconfident if $x-c>0$ (and under-confident if $x-c<0$). Let us translate to confidence in swimming ability. Intuitively, overconfidence in swimming ability might mean two things: a) that we believe to be able to swim 500 metres when in fact we are only able to swim 100 metres; b) the mean subjective probability assigned to the correctness of a proposition like "I can swim back to the shore" minus the proportion of successful cases being positive. b) is the intended interpretation in this context. The apparent phenomenon of overconfidence has important philosophical and psychological explanations and suggested remedies. One of the reasons falsificationism is attractive is that we believe in Bacon's judgement that we are victims of selective retrieval of supporting evidence. Nowadays, the idea of a "cognitive overconfidence bias" is considered an established fact by many.[24] Moreover, research has documented empirically "the hard-easy

[24] See, for instance, Griffin and Tversky (1992).

effect"[25] which occurs when the degree of overconfidence increases with the difficulty of the task.

However, as can easily be seen in the following graph high success rates effectively bar the possibility of overconfidence:[26]



That overconfidence is a property of difficult tasks and underconfidence of easy ones is at least partly an artefact of the criterion employed.[27] The effect is especially interesting in our example. Since the number of failed attempts to swim back to the shore cannot be far from 1 (i.e. the proportion of successful cases including all swims is normally close to 1) any

[25] Griffin and Tversky (1992).

[26] Adapted from Juslin et al (2000).

[27] Compare Juslin et al (2000).

explanation of drowning accidents building on the psychologist's concept of overconfidence must ensure that the reference class does not include the easy cases. Let us therefore restrict the reference class in some suitable way to include only cases similar to the unfortunate and harder one. As a result, the frequency generated by the single swimmer won't do because it contains too few instances to make possible any precision in verdicts of confidence. So the frequency has to be expanded in some way, either by taking in other swimmers sufficiently similar to the actual one or by employing hypothetical or simulated cases.

Even such a measure of overconfidence in swimming ability has its problems. Is it OK to apply a frequentistic criterion (the hit or success rate) to a nonfrequentistic or subjective concept of probability? Conceptual problems like these make it questionable to speak of the overconfidence finding as a mathematical bias.[28] But at least for some theories, the conceptual problem shouldn't be a practical one since their point of origin is that we are calibrated in such a way. For the Bayesian, for instance, the criterion should make sense. Nevertheless, overconfidence is often formulated in more cautious terms that do not require a perfect match between probability and frequency: "A person is said to exhibit overconfidence if she overestimates the probability of her favored hypothesis. The appropriate probability estimate may be determined empirically (e.g., by a person's hit rate) or derived from an appropriate model."[29]

---

[28] See especially Gigerenzer et al (1991).

[29] Griffin and Tversky (1992), 412, footnote 1.

Elster's use of confidence in his mechanistic version of The best swimmers drown fits nicely into this tradition of cognitive psychologists' gathering of a massive body of experimental knowledge that seems to show what Bacon suspected, that humans are naturally prone to overestimate their knowledge. This observation is what is needed to suggest that confidence in Elster's mechanistic explanation is primarily related to another kind of risk than ability is. This causal process in Elster's model leads to *epistemic* risk.

Those who want to extract my constructive critique of Elster should skip the following paragraphs and head for section 8. I find the relations between confidence and epistemic risk much too interesting to follow those readers at once. Advocates of epistemic risk claim that unnecessary risk taking may occur if one does not pay attention to what one does not know.[30] This seems undoubtedly true. Does this truth translate to issues of confidence?—for instance, to the claim that in order to avoid unnecessary risk taking one should consider not only which beliefs one has but one's confidence in them? Let us look at how the idea of epistemic risk is actually introduced in two different cases:

Gärdenfors and Sahlin (1982)

This paper contains a moderately famous setup concerning Miss Julie who is invited to bet on the outcome of three different tennis matches. In Match A, Julie is well-informed about the two players. She predicts that the match will be very even. In Match B, Julie knows nothing about the players. In

[30] Sahlin and Persson (1994), and Gärdenfors and Sahlin (1982/1988)

Match C, Julie has overheard that one of the players is much better than the other but—since she didn't hear which of the players was better—otherwise she is in the same position as in Match B. Now, if Julie is pressed to evaluate the probabilities she would say that in all three matches, given the information she has, each of the players has a 50% chance of winning.

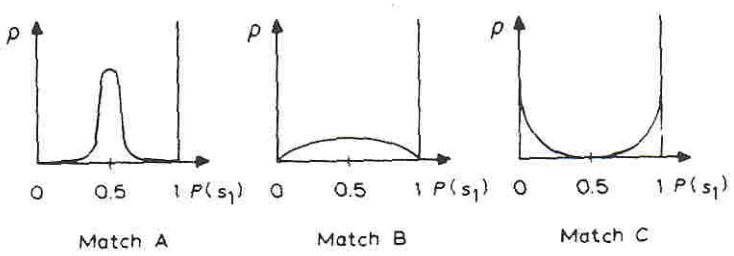
> It seems, however, perfectly rational if Miss Julie decides to bet on match A, but not on B or C, for the reason that a bet on match A is more *reliable* than a bet on the others. Furthermore she would be very suspicious of anyone offering her a bet at equal odds on match C, even if she could decide for herself which player to back. [31]

The example is intended as a refutation of a strict Bayesian approach, according to which the decision maker's knowledge in a given situation can be represented by a unique subjective probability measure defined over the possible states of the world. Situations such as in the example show that the amount and quality of information the decision maker has is an additional factor of importance. In order to account for this idea Gärdenfors and Sahlin introduce the notion of *epistemic reliability*. It comes in degrees. The upper bound represents the case when the decision maker has complete information. The lower bound represents the case when the decision maker has no information at all. First, following Levi (1974/1982), a state of belief is represented by a class of probability measures. Then, a measure p of the epistemic reliability of the probability measures is introduced.

---

[31] Gärdenfors and Sahlin (1982/1988), p. 314.

Technical details are of no importance here. A graph of the three matches above picturing the state of belief that player 1 wins and its epistemic reliability is enough for purposes of illustration:



How does this connect to epistemic risk? The idea is that it is reasonable to assume that Miss Julie perceives a greater risk in betting on match B and C, than in match A. "An agent who takes all epistemically possible measures into consideration takes no 'epistemic' risk at all"[32], but allowing only the unique probability 0.5 in B and C leaves many of the epistemically possible measures unconsidered.

The transition from epistemic reliability to epistemic risk taking is thus obtained via decisions of which probability distributions are accepted as *serious possibilities*. "Deciding to consider some distributions in P as not being serious possibilities means that *one takes a risk*, an epistemic risk.[33]

---

[32] Gärdenfors and Sahlin (1982), p. 323, footnote 16.

[33] Gärdenfors and Sahlin (1983), p. 242.

Gärdenfors and Sahlin (1983)

An enlightening comparison between confidence studies of the above type and epistemic reliability occurs in this succeeding paper:

> If an almanac question is employed, the (event) probability of an alternative is either 1 or 0. Absinthe is either a precious stone or it is not, so P(a) = 1 or P(a) = 0. Or, in other words, the almanac propositions are closed propositions, the truth values of which are already determined. Thus, subjects' assessments of how confident they are in their answer will not be a primary probability assessment, but a second-order assessment. In terms of the present theory, subjects make an assessment of p(P(a) = 1), i.e. they estimate the epistemic reliability of P(a) = 1.[34]

Here, then, a rather intimate connection between confidence and epistemic reliability is foreshadowed.

Summary

In sum: epistemic risk is introduced via the notion of epistemic reliability and serious possibilities. Epistemic risk taking occurs when only one of two beliefs of similar epistemic reliability is regarded a serious possibility. Moreover, many of the examples where epistemic risk taking are introduced are fairly intuitive. The examples build on situations where the decision not to regard a certain probability distribution as a serious possibility is a reconstruction of the actual situation. And so is often the

---

[34] Gärdenfors and Sahlin (1983), p. 246.

acknowledged probability (Miss Julie *is pressed to evaluate* the probabilities.)[35] Perceived confidence does much of the actual work—the same work in fact as perceived epistemic reliability of the acknowledged serious possibility is supposed to do according to the theory. Notice: this is not a critique. It is clear that this tradition has the theoretical resources to approach the cases differently. Epistemic risk is independently defined. Epistemic risk can be more analytic where confidence is always more of a valuation. But still, in the examples and probably in the empirical world in general there is a pretty important correlation between confidence and difference in degree of epistemic reliability.

7. The importance of calibration

So far, in the comparison of epistemic risk and confidence calibrated decision making has been in focus. If one is calibrated, upon reflection perhaps, high confidence seems to imply low epistemic risk. But this important correlation builds on the calibrated decision maker.

Can this correlation be expanded to cover the non-calibrated decision maker too? The best swimmers drown illustrates a case where unnecessary risk occurs because the swimmer considers and acts on his or her confidence (which happens to be poorly calibrated in the overconfidence direction). Confidence is high and perceived epistemic risk is accordingly low. Someone only interested in putting forward a descriptive theory need not be worried. That epistemic risk is perceived to be low will function well

[35] Gärdenfors and Sahlin (1982), p. 314.

in an explanation of the actual behaviour. But for someone interested in promoting epistemic risk as a normative theory the non-calibrated decision making constitutes a possible difficulty. Taking notice of epistemic risks via degree of confidence seems to lead us wrong in cases of overconfidence. And if the phenomenon of overconfidence is massive, as some psychologists think, acknowledging epistemic risk of this kind in decision making might make more harm than good. I will touch briefly on two escapes. Either the overconfidence phenomenon has to be shown to be merely chimerical or restricted at least to fewer situations than is commonly recognised, or degree of epistemic reliability should not be measured by degree of confidence. I think that some things can be said in favour of both.

When overconfidence amongst drowned male swimmers has been discussed in Swedish media this summer, two explanations have been tried out.

1) These men used to have the ability it takes to survive open water conditions, but over the years the ability was gradually lost without them noticing the change. Five years ago they easily swam back to the shore, but this time they couldn't.

2) These men had the ability but only in indoors conditions. Maybe they used to swim 300 metres every week in the bathhouse, but swimming outdoors is sometimes very different.

The inference from 1) or 2) to "I know that I can swim into the shore if I fall into the water" is understandable but as a matter of fact mistaken, the reason being that 1) and 2) are not sufficiently similar to this situation. *If*

the reference class had been another, including the easier tasks, or if swimming in open water had been more similar to indoors conditions, then the inference from 2) would not have created overconfidence. *If* the person had been the man he used to be then nor would the inference from 1) have created overconfidence. Gigerenzer and colleagues have put forward an ecological model which is based on the assumption that people are good judges of the reliability of their knowledge, "provided that the knowledge is representatively sampled from a specified reference class."[36] Their model goes some way towards delimiting the number of cases of overconfidence but it cannot be used to explain the overconfidence in the case of the best swimmers. These swimmers are not exposed to the experiments of others but entirely free in their choice of reference class. Still a somewhat similar kind of process might be going on when the best swimmers drown.

It is quite interesting actually to follow the debate on what kind of swim training our schools should offer. This debate shows that even the experts disagree on what characterises the dangerous swimming cases. For instance in a recent article, Anders Wernesten at Swedish Life Saving Society is disappointed that the Swedish National Agency for Education now suggests that every student in fifth grade should be able to swim 200 metres. He thinks that 50 of the 200 metres should be backstroke. "This is very important because you are able to rest on your back". The Swedish National Agency for Education does not agree: "The way you swim is of

[36] Gigerenzer et al (1991), p. 506.

minor importance. What it is all about is to be able to survive an emergency situation in water", responds the person in charge.[37]

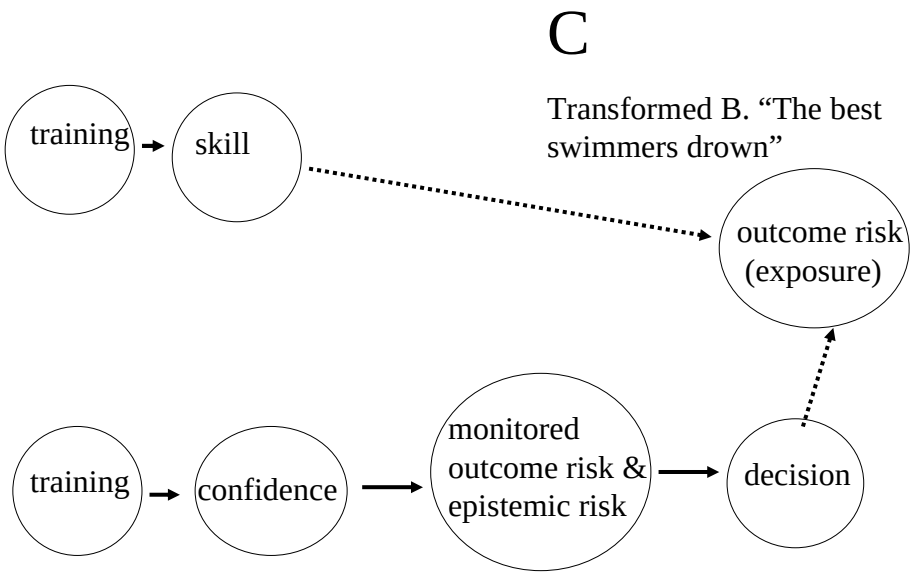
It seems to me that in order to be a good normative theory[38] an account of epistemic risk has to be able to accommodate exactly the kind of mistakes displayed in the inferences from 1) and 2) to "I know that I can swim into the shore if I fall into the water", as well as the uncertainties manifested in the debate above. Such inferences and decisions create considerable epistemic risks that go unnoticed by the agent and thus lead to unnecessary or unwarranted risk taking. Hence, epistemic risk should not be linked too closely to degree of confidence. Again, this is not a critique of Sahlin and colleagues but it points to the need for other kinds of example introducing epistemic risk, for instance one that makes the difference between perceived and objective epistemic risk clearly visible.

8. Improving the picture

Let us now put some of the pieces together. Instead of figure B the mechanistic picture should look something like C:

[37] "Skärpta krav på barns simkunskaper", DN.se, September 1 2006.

[38] "Normative" can be used in different ways too, as Niklas Vareman has shown me. Here I interpret normativity in objective terms.



Assuming for the sake of the argument that the added pieces are correct causal additions to the picture, a development such as this one ought to secure explanatory progress in the process from B to C. In general it is assumed that making an explanation more fine-grained adds to its explanatory value (as long as the explanation is still understandable). And mechanistic explanations in particular should benefit from this kind of development. It is true that “mechanistic explanation” can mean a lot of things, but here we get a clearer picture of how the risk taking effect is produced, how the mechanism producing this result is structured, and how to fill in some of the missing links between cause and effect. A

development such as the transition from B to C strengthens the mechanistic explanation along the three dimensions of *production*, *complex systems*, and *the hidden tie*.

9. Elster's problems

It is by no means clear, however, that mechanistic explanations as Elster understands them gain anything from a move from B to C. To the contrary, development constitutes a constant threat to mechanistic explanations of his kind. The reason is that there is a requirement on all type-B mechanisms that they have indeterminate consequences. But this can of course not be guaranteed if we admit a process where the mechanistic picture is filled in. Two interpretations of the explanatory situation in C illustrate breakdown-scenarios.

The first mechanistic breakdown would occur if the picture in C is taken to suggest that the two causal processes are about completely different aspects of the situation. There is one dimension of exposure and objective outcome risk (the process above), and one dimension of decision, epistemic risk and monitored outcome risks (the process below). We simply give up on finding an explanation linking the two – remember Hempel's *Aspects of Scientific Explanation* where he convincingly argues that if we want a scientific explanation we cannot allow the explanandum to be too complex.[39] The dotted arrows in C might be taken to show a situation where two distinct processes have a causal influence on different aspects of

[39] Hempel (1970), p. 422.

the risk exposure and where risk exposure as a concrete event is not within either explanation's scoop. The consequence of such a division, according to Elster's understanding, must be that the mechanism disappears and brings the complete mechanistic explanation along. This is rather counterintuitive. True, the explananda have changed with this new direction of research but the component explanations should still exist – and should have been considerably strengthened. But then on Elster's view they have no mechanistic foundation any longer. This is the *structural change* objection.

The second mechanistic breakdown occurs if we given the additions in C are in a much better position to actually decide how someone's training affect his or her (outcome) risk exposure. But according to Elster the consequences of a type-B mechanism must be indeterminate. So it cannot be a type-B mechanism, and hence no mechanism any longer. This is the *known consequence* objection.

Elster's account of mechanisms faces two rather severe difficulties that other mechanistic approaches do not. The truth is that this creates an even more problematic situation when it comes to our understanding of explanation. Recall Elster's conviction that we need an explanatory middle level for cases where covering laws are rare. In the figure below, (0) marks the explanatory situation before mechanisms and (B) a somewhat optimistic situation with Elster-style mechanisms. Then, between (B) and (C), a mechanistic breakdown occurs. Let us assume that one result of this breakdown is that the number of covering law-explanations increases. But it cannot be taken for granted that all chartered causal processes in the previous mechanisms will qualify as covering law-explanations. Structural changes as in the first mechanistic breakdown should not result in covering

law-explanations (indeed if there were laws governing the component processes, there would probably have been a resultant law governing the mechanism to begin with, and thus no indeterminate consequences.). And I cannot find a convincing reason why known consequences in general should result in covering law-explanations. It seems that this will only be the case where some previously unknown law or initial condition for an already identified law is added to the mechanistic picture. But as far as I understand Elster's mechanisms there is no assumption that the previously acknowledged components of his mechanistic explanations in (B) (the components to which new pieces were added) were of this kind. Furthermore, laws usually have a number of properties which (the components of) Elster-style mechanisms need not have. Apperently, there is a substantial mismatch between laws and Elster-style mechanisms which will infect the possibilities to go from mechanisms to laws when a little more causal knowledge is added:

Laws, on the one hand, are generalisations of wide scope that apply to many different kinds of systems. They have no (or few) exceptions. They cannot easily be disturbed. Mechanisms, on the other hand, might not generate suitable generalisations at all. They can be very specific, unevenly distributed, and fragile.[40] This means that the development of the causal picture, the causal epistemic progress, we see in the transition from B to C, might not lead to the discovery of laws or the completion of initial conditions for such laws.

---

[40] Compare Persson (1999).

Let us be fair to Elster. It is entirely possible that some transitions from B to C increase the number of covering law-explanations, and it is at least as clear that some will not. Let us finally focus on the latter cases, where we have increased our causal knowledge, lost our mechanistic explanation but not achieved a covering law-explanation in return. This causal knowledge belongs to yet another level than description, mechanistic explanation, or covering law. From the explanatory perspective of ordering that we have assumed, such causal knowledge is located on a level between description and mechanistic explanation. In sum, stage (C) is characterised by a situation where we have introduced a new non-explanatory level and increased the number of covering law-explanations.

(0)	**(B)**	**(C)**
Explanation by law	Explanation by law	Explanation by law
No explanation: But possibly causation without laws	Explanation by mechanism	Explanation by mechanism
		No explanation: causation without laws
Description	Description	Description

I think these consequences should be strongly counterintuitive according to most of us. They should be especially damaging for anyone who tries to build a causal understanding of explanation.

References

CBC News Online July 19 (2005). Indepth consumer tips: Kids and water safety: Myths and stats.
http://www.cbc.ca/news/background/consumertips/water-safety.html

Darden, L. (2002). Strategies for Discovering Mechanisms. *Philosophy of Science* 69: 354-365.

DN.se July 9 (2006).

DN.se August 23 (2006).

DN.se September 1 (2006).

Drunkningsolyckor 2005. Svenska Livräddningssällskapet.

Elster, J. (1998). A plea for mechanisms. *Social mechanisms*. Hedström and Swedberg (eds.). Cambridge University Press: 45-73.

Elster, J. (1999). *Alchemies of the mind*. Cambridge University Press.

Elster, J. (forthcoming 2007). *Explaining social behavior*. Cambridge University Press.

European Child Safety Alliance. Be water wise—water safety and drowning prevention checklist.
http://www.childsafetyeurope.org/csi/ecsa.nsf/index/watersafetycampaign/$file/WWchecklist.pdf

Gigerenzer, G., Hoffrage, U., and Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review* 98(4): 506-528.

Glennan, S. (2002). Rethinking Mechanistic Explanation. *Philosophy of Science* 69: 342-353.

Griffin, D. and Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology* 24: 411-435.

Gärdenfors, P. and Sahlin, N.-E. (1982/1988). Unreliable probabilities, risk taking, and decision making. *Decision, probability, and utility*. Gärdenfors and Sahlin (eds.). Cambridge University Press: 313-334.

Gärdenfors, P. and Sahlin, N.-E. (1983). Decision making with unreliable probabilities. *British Journal of Mathematical and Statistical Psychology* 36: 240-251.

Hempel, C. G. (1970). *Aspects of Scientific Explanation*. Oxford, Glencoe.

Juslin, P., Winman, A., and Olsson, H. (2000). Naïve empiricism and dogmatism in confidence research: a critical examination of the hard-easy effect. *Psychological Review* 107(2): 384-396.

Levi, I. (1974/1982). On indeterminate probabilities. *Decision, probability, and utility*. Gärdenfors and Sahlin (eds.). Cambridge University Press: 287-312.

Mackie, J. L. (1974). *The cement of the universe*. Oxford, Clarendon Press.

Persson, J. (1999). On the determinables of explanatory mechanisms. *Synthese* 120(1): 77-87.

Railton, P. (1978). A deductive-nomological model of probabilistic explanation. *Philosophy of Science* 45: 206-226.

Sahlin, N.-E. (1993). On higher order beliefs. *Philosophy of probability*. Dubucs (ed.). Kluwer academic publishers: 13-34.

Sahlin, N.-E. and Persson, J. (1994), Epistemic risk: The significance of knowing what one does not know. *Future risks and risk management*. Brehmer and Sahlin (eds.). Kluwer Academic Publishers: 37-62.

Salmon, W. (1997). Causality and explanation. Oxford: Oxford University Press.

Saltzman, S. (2005). Les Liaisons Dangereuses. *Theatre scene.net* 07/13/2005

Siegel, N. (2005). War of the Boudoirs. *The New York Times* July 17.

Svenska Dagbladet July 31 2006.