The Risk Homeostasis Theory: Accept, reject or modify?  
- An Opposition to Gerald Wilde’s RHT

Truls Vaa, Institute of Transport Economics, Gaustadalléen 21, NO-0349 Oslo, Norway,  
+47 2257 3825, tva@toi.no

The impact of Wilde’s Risk Homeostasis Theory

The impact of Wilde’s Risk Homeostasis Theory (RHT) since it was first launched by Wilde in 1982 has been tremendous and still is. It is impossible to draw up a complete map of where and when this impact has taken place, but what I can do is to sketch the impact Wilde has had on the research environment I know of and where I have worked since 1988 - The Institute of Transport Economics (TØI). Here are some basics of the important impacts of Wilde’s RHT at TØI:

- Risk compensation has been a standard concept and an integral part of discussions on driver behaviour and traffic safety measures for more than 20 years
- 1993-1997: A Strategic Institute Program (SIP) on Risk Compensation financed by the Norwegian Research Council and resulted in two important papers
  - 1) An investigation of behavioural adaptation to airbags and antilock brakes among taxi drivers (Sagberg et al, 1997)
  - 2) Risk compensation and car drivers’ behavioural adaptation to the installation of road lighting (Assum et al., 1999).
- 1998-2003 A Strategic Institute Program (SIP) on Driver Behaviour Models financed by the Norwegian Research Council, Norwegian Public Roads Administration, Ministry of Transport and TØI
- 2007: Modelling driver behaviour on basis of emotions and feelings: Intelligent transport systems and behavioural adaptation (Vaa, 2007)
- 2007: Explaining and proposing four hypotheses on the prediction of effects of ITS (Vaa et al, 2007)

Some assertions

In a paper presented at the ICTCT-workshop in Caserta 2001, I put forward some assertions that may still be valid (Vaa, 2001a):

- Disagreement prevails: There is no consensus concerning driver behaviour models, no agreement has been achieved, no break-through or “GUT” (Great Unified Theory) has emerged. Existing models address different aspects. Each of the aspects may be significant, but there is no general integration that links the diversity of aspects into a complete and all-embracing model.

- Attempts of modelling information processing, decision-making, thinking and feeling has been “too much of a black box”. The main cause: A lack of thorough and comprehensive understanding of human cognition and emotion, i.e. how we, the drivers, think and feel. Recent achievements in cognitive psychology and neurobiology have not been applied on prevailing driver behaviour models.

- The Taylor statement: “Driving is a self-paced task governed by the tension/anxiety the driver wishes to tolerate” may be more in line with more recent achievements in neurobiology – as in Bechara et al (1997) - than any other of the driver behaviour models that exist.

- No deep understanding of risk compensation will emerge unless recent developments in neuroscience are integrated in the modelling of driver behaviour.

The most fundamental of these assertions is the last one about risk compensation and the main problems are: 1) Risk compensation exist, and 2) No satisfactory explanation has been provided.

Wilde’s model

One of the basic starting points of Wilde’s RHT is in control theory/cybernetics. Wilde states very clearly that his RHT-model may be compared to a heating system that regulates the in-door temperature of a house (Wilde, 2001).

Figure 1: Homeostatic model relating house temperature to heating system activity (from Wilde, 1988).
The homeostatic model relates house temperature to heating system activity and vice versa: Relating heating system activity to house temperature, with the set-point (target) temperature as the controlling variable (Wilde, 2001; Wilde, 1988). Wilde converts his model of regulating the house temperature “box-by-box” in developing his model of how the RHT works (Wilde, 2001):

![Diagram of Wilde's Model of Risk Homeostasis](Wilde, 1982).

Wilde describes the model as a

“Homeostatic model relating the accident rate per head of population in a jurisdiction to the level of caution in road-user behaviour and vice versa, with the average target level of risk as the controlling variable” (Wilde, 2001 – page 33)

Core issues and problems

There are three main issues here, which I regard as a core of an opposition against Wilde’s RHT:

1. The notion of a comparator or summing point: Where – in the organism – is this “comparator”? Is it regarded as a conscious entity or activity, an unconscious one, or both?

2. The model description above refers to “…the accident rate per head of population in a jurisdiction ….”. By this notion it becomes unclear whether the RHT is model that operates on a disaggregate (individual) or aggregate level or perhaps both, although Wilde says explicitly:

   “… please note that Figure [above] does not refer to a single individual, but to all road users in a given jurisdiction such as a city, township, county, province or nation” (Wilde 2001, page 31)

I must rebut Wilde’s statement, because, as the description indicates, this indeed must be a cognitive operation to be done – and felt - by the individual: He/she must have a mental concept of a jurisdiction, know the aggregate number of
accidents in this jurisdiction, calculate the accident rate per head, adapt to this accident rate by contributing his/hers share to the aggregate number of accidents so that homeostasis in this jurisdiction is maintained. I assert that this is impossible for an individual to do and the hypothesis should be easily falsified. The alternative, which might be what Wilde suggest, is that an individual might read about, hear, discuss, that the accident rate in his/her jurisdiction is, changing and then adapt to this accident rate by his/her individual accident contribution by increasing or reducing it so that homeostasis is restored or maintained. If the homeostasis mechanism does not act through this individual contribution, then the whole population of a jurisdiction must act *aggregately*, “in the direction of homeostasis” as some kind of a tremendous swarm as were they (we!) starlings. That just do not make sense!

3. Finally then, is the *target* concept which indeed is a core concept in RHT. The point here is how this target is understood. Inherent in an understanding lies an implicit conception of a target risk which must be some kind of accident number or accident per capita in a jurisdiction around which the homeostasis mechanism is oscillating. However, the target concept might also be understood and defined differently and rephrased in a direction of risk compensation or the more broad behavioural adaptation – i.e. not as strict and exact as with risk *homeostasis*.

**Risk comparing or risk monitoring?**

The question of whether risk is compared or monitored is a basic issue to be dealt with and two simple historical lines of development might be drawn up:

![Figure 3: Historical lines of development of some driver behaviour models](image)

I see two historical lines which both are based Gibson and Crook (1938) and Taylor (1964): The one founded by Wilde (1982) and the one founded by Näätänen and Summala (1974). The central issue is then the question of how risk is appraised and how processes linked to risk appraisal should be modelled. The concept of a *comparator* produces an image of something being compared (consciously) and that behaviour are
triggered by differences between images or ‘inner scenarios’ (Vaa and Bjørnskau, 2002). The appraisal of images/inner scenarios means that specific comparisons between these scenarios are made, especially in contexts which demand (conscious) choices between alternatives. The concept of a comparator is a basic element in Wilde’s RHT, but further development of the RHT is considered to be a deadlock, unless it is rephrased. Note also that the target risk in this context is a number > 0.

Monitor, however, implies something being monitored more or less continuously. The main task of a monitor is, as the word says, to monitor the organism and the situation in which it operates, identify unpleasantness and danger in order to avoid it, and/or to achieve a better condition among those scenarios that are available as alternatives. The basic objective of the monitor is to secure or increase the probability of survival. In monitoring there is not necessarily anything being compared and no standpoint is taken towards whether consciousness is involved. A monitor functions universally rather than specific and would not be viewed as a focal point as with risk comparing. The concept of risk monitoring was introduced by Näätänen and Summala and is a key concept of their “Zero-Risk Model” (1974). Note that target risk in this context is zero.

The development axis along the monitor line of figure 3 indicates what may be achieved if one base the elaboration of a driver behaviour model on the monitor concept. Damasio (1994) and Bechara et al (1997) are neuroscientific contributions that may serve as inputs to the development of a driver behaviour model based on risk monitoring. While Bechara et al confirm Taylor (1964) and the role of Skin Conductance Response as a guide in monitoring risk, Damasio simply states axiomatically that

- *Man’s deepest motive is survival*

It follows from this axiom:

- **We must have an organ, a risk monitor for detecting dangers that threaten survival**

Assertion (1):

- **the body is the risk monitor**

Assertion (2):

- **Emotions and feelings are the tools enabling the organism to monitor the risks of the environment in which the organism operates**

Damasio introduces what he labels an “unorthodox definition” of emotions and feelings by limiting emotions to unconscious/automated processes – as with the schemes - and feelings to conscious processes – as with conscious appraisals of inner scenarios/images (Damasio, 1994).

**Further comments on theoretical issues and problems with the RHT**

It has been heavily debated whether Wilde’s RHT is a theory that can be tested and it has been accused for being circular:

- If a measure does not reduce the number of accidents, the theory is confirmed (homeostasis is retained)
- If it is found that the number of accidents is reduced, it can be explained by stating that the target level of risk has been reduced
As a consequence, no result will lead to a rejection of RHT

Further, the issue of whether the RHT is a theory of individual behaviour or “aggregate jurisdictions” also contributes to problems with scientifically testing. Consider the following statement:

"…. It should be noted that RHT is a set of interrelated hypotheses developed to explain the accident rate of large numbers (often millions) of socially-interacting road users over a considerable length of time. Breakdown of control causes some people to have accidents. These accidents subsequently serve as danger signals to others and help the majority to avoid them" (Wilde, 2001 - page 150)

On the one hand, Wilde speaks of “…. the accidents rate of large numbers (often millions)….” – and on the other hand, “Breakdown of control causes some people to have accidents”. By this statement the aggregate (jurisdictions) and the disaggregate (the individuals) levels are so intertwined that your really cannot separate between the two, or sort out on which level you should test your hypotheses. One of the mechanisms that Wilde suggests as bridging the gap from the aggregate number of accidents to the individual accident level is by “witnessing accidents, reading newspapers (about accidents) and discuss, a process which take some time and results in the “lagged feedback” in Wilde’s model. Again, I question such a mechanism, because:

- Accidents are rare,
- We do read about them, but do we really talk about them?
- If we have one accident, we don’t like to talk about it and we don’t like to think too much about accidents when we are driving.

May I add: I don’t think we should imagine concrete accidents and injuries too much, it is adequate to think dangers and risks, yes, but to make too much of it may make us impaired when driving. In fact, there is some empirical evidence that being too frightened in traffic may impair driving (Näyhä, 2002).

**Target risk > 0 or target risk = 0?**

One basic disagreement with Wilde, which is illustrated by figure 3, concerns the value of the target risk as imagined by the individual human being: Do we imagine it as zero or greater than zero? Consider the following statement:

"….. all behaviour is risk-taking behaviour, regardless of whether this is consciously realized by the acting person or not. It is obvious, too, that the challenge of life is not to eliminate risks. ”Zero risk” is not a meaningful option, since it can only exist in the absence of behaviour – after death, in other words” (Wilde 2001, page 151)

In this I disagree fundamentally with Wilde and it illustrates very concretely the difference between Wilde’s conception of risk and the conception of risk in other models. And again I think the confusion is brought about by not being absolutely clear about
which level one operates. It is obvious that a society encounters and will encounter accidents within its jurisdiction within a given time span but that is of course not the same as saying that an individual will encounter an accident within the same time span. In fact, it is more normal that a driver will not experience an accident with personal injury in his/her lifetime than he or she will. A calculation based on Norwegian accident statistics estimates the individual involvement in an accident with personal injuries to, on the average, be every once in 390 years, that means that you must have a group of some 6 drivers driving ca. 14,000 km a year for 65 years, i.e. from they are 18 to they are 83 years of age, before one of them, on the average, will experience one personal injury accident (Vaa, 2003). And in this single accident, the probability that it will result only in some minor injuries is 70-80%. In fact, this is a very good empirical illustration of how skilled the individual driver is in monitoring and avoid dangers in traffic. Again: it is normal to drive a whole lifetime without experiencing a road traffic accident with personal injury.

Wilde illustrates his conception of risk by utility functions as in the following figure:

Figure 4: Theoretical representation of road users as net benefit maximizers (Wilde, 2001)

In the above graph \( y_3 = y_1 - y_2 \). Through this theoretical representation of benefits, Wilde characterizes road users as “risk optimizers” because they:

“…… choose an amount and manner of mobility such that the associated level of subjective risk corresponds with the point at which the expected net benefit is maximal.” (Wilde 2001, page 35).

The target – as understood and defined by Wilde – may by chosen so that the value of “level of exposure to risk” result in an optimum where the utility is at its maximum. Then concepts which Wilde applies here are taken from economy and it is a question whether these concepts of “utility”, “expected loss/gain” and “optimal target level” on the whole are applicable when it comes to how drivers deal – i.e. psychologically speaking – with risks in the road system. In my view, the utility theory and adherent concepts are not
applicable in this context as it rules out fundamental topics as information processing and unconscious and conscious routes to decision-making in dealing with the risks and dangers of every-day road traffic. One cannot simply just rule them out and leave them uncommented as they do not exist. The alternative is of course to go deep into the subject matter as for example Reason has managed to do (1990), by modelling what he calls “a fallible machine” of information processing and decision-making (figure 5).

![Figure 5: Reason's model of information processing and decision-making (Reason, 1990)](image)

Reason’s “fallible machine”-model illustrates the complexity of interrelationships between entities of memory, storages of information and types of knowledge. Inherent in Reason’s model is the building of experience through schemes – i.e. the encounters with myriads of different scenarios of everyday life, in road traffic as well as outside, that we have to face and deal with as they come. One important point here is the process of automation, probably initially starting as a conscious process regarding how to solve a problem, appraise it, try it out, fail perhaps, before the “best solution” is found, and as time goes on, transfer the conscious appraisals to automated schemes and acts without thinking consciously about the problem any more, the experience is “in the body” – it is there when needed, also in terms of dealing with risks and handling dangers (Vaa, 2007).

Finally then, to complete the discussion on the value of the target risk, a completely different type of understanding is suggested as opposed to Wilde’s utility model. Consider the following simple model:
The graph of figure 6 is a deduction from Näätänen and Summala’s “Zero-Risk Model” (1974) and it is put forward to illustrate the following:

1. The perceived (subjective) risk of an accident = 0, for all values of $x < x_1$.

2. For all values of $x > x_1$ the perceived/subjective accident risk > 0.

Again: For all values of $x < x_1$ the perceived/subjective accident risk is zero. Every $x < x_1$ realises the feeling of a subjective accident risk which is zero. Why then do drivers stop at $x_1$ as the chosen driving speed? Why not choose any other of the speeds $< x_1$? Why exactly $x_1$?

Näätänen and Summala do not state or answer this question directly. But let us see what happens if we loosen Wilde’s rigid presupposition about estimating the target risk level as a certain number $> 0$ that the driver seeks to achieve. Let us suppose that this target is of another nature. Let us suppose instead that drivers are searching a certain feeling, a certain way of driving that suits him or her well - “a best feeling” - which is realised in the organism by his or her choice of $x_1$. Then the target should not be regarded as a number, but rather as a certain kind of feeling – a target feeling. A mere fulfilment of a “zero-risk” is then not enough, there has to be added another dimension to it as well. A dimension, or an experience, that is achieved at the “exact speed of $x_1$ “, but not at speeds lower than $x_1$. Then, the following assertion can be stated:

**Assertion:** In addition to avoid accidents, drivers seek a certain “target feeling”. This feeling is not the same in all drivers, all drivers have a unique target, which is not necessarily experienced consciously. Targets that drivers seek are defined and characterised by an emotional dimension. (Vaa, 2001a)

This is exactly what I would characterise as Wilde’s contribution by his RHT, the introduction of **the target** – but the RHT has to rephrased: It is not to be understood as a number or some probability fraction of a risk of accidents, but as a feeling – a best feeling which drivers are seeking and which is possible to realise in a given context in road traffic.
There is “a best feeling” also inherent in Wilde’s RHT, - which is phrased in economic/utility terms as:

- Benefits/costs expected from risky/cautious behaviour (+/-)

Applying utility terms is, however, not sufficient because it does not fully grasp the psychological dimension of emotions – emotions is not stated explicitly in Wilde’s utility terminology. This can be illustrated by stating several “emotional candidates” that would correspond better to the variety of personality traits that drivers exhibit. Other candidates of best feelings that drivers may seek are then:

- "Arousal” and being vigilant, attentive, aware, focused
- Sensation (seeking)
- Pleasure
- Security, minimizing workload
- Avoid violations (always behaving correctly)
- Non-compliance

I shall not go into much detail about learning theory and reinforcers of behaviour, but just mention it: All the above emotional dimensions may be classified as reinforcing stimuli – or $S^R$ - in learning theory terminology.

**Personality traits, attitudes and lifestyle**

Personality and especially personality traits, has entered the field of traffic safety research in recent years, but the personality model used by Wilde is quite limited:

"In general, it has been found that correlations between personal characteristics and accident record, when they exist at all, are weak and often statistically unreliable” (Wilde 2001, page 170)

"… no general risk-taking personality trait emerges from the research studies” (Wilde 2001, page 170)

"If there is one particular habit that is often believed to be associated with risk-taking personality traits, it must be gambling, especially pathological or compulsive gambling. In everyday language,"to gamble” is almost synonymous with ”to take a risk” (Wilde 2001, page 171)

"…the many studies undertaken in order to identify traits that might characterize these people have failed to indicate that gamblers are more-than-average risk takers …..” (Wilde 2001, page 171)

"…there would seem little hope for finding any marked relationship between accident involvement and personality” (Wilde 2001, , page 172)

All these statements can be rebutted by taking into account what Ulleberg documents in the relationship between personality traits and accidents (2002). Ulleberg applies the “BIG-5” test-battery of personality traits where the basic factors are:
• **Extraversion:** social, active, seeking adventures and thrills, dominating
• **Neuroticism:** Anxious, hostile, depressed, variable mood, impulsive, vulnerable
• **Conscientiousness:** Conscientious, self-discipline, organized
• **Agreeableness:** Empathy, relying on others, helping others, conform, pleasing
• **Openness:** Imaginative, creative, open for new ideas

What Ulleberg manages to do is to link attitudes, motives, behaviours and accidents with personality traits. By cluster analysis he discern between six subgroups of drivers by also indicatively labelling them by a “key characteristic” (2002).

1. “Considerates” (15%): Balanced, calm, low on anxiety/aggression: considerate and caretaking, smooth interaction, avoiding conflicts, respecting law, low on anxiety and aggression, slightly more women than men, (accident risk below average)
2. “Socially deviants” (15%): Normless, irresponsible, low on altruism, egoistic, (very) self-confident, low on anxiety, sensation-seekers, non-compliant, low on consideration, create conflicts, low on understanding the risks of behaviour. 80% men, accident risk above average
3. “Anxious” (15%): High on anxiety, high on altruism, low on stimuli-seeking, insecure, avoiding conflicts and workload, 84% women, accident risk below average
4. “Considerate sensation-seekers” (22%): High on sensation-seeking, high on altruism, moderate on normlessness, average accident risk
5. “Aggressive” (15%): High on aggression, anxiety and driving anger, irritable, low on altruism, low on skills, low on consideration, lower on self-confidence than the “social deviants”, high on sensation-seeking, hostile, making conflicts, unsolved conflicts. 57% men, accident risk above average
6. “Adaptable, but egoistic” (18%): Moderate all-over, low on sensation-seeking, altruism, and consideration, strategic, selfish, self-control, avoiding conflicts. (“economic man”?), 57% men, average accident risk

By this description and characterisation of driver subgroups it becomes clear that the frequency of accidents must be associated with personality traits and probably also explained by personality constitution and emotional problems as might be seen in the driver subgroups labelled “Socially deviant” and “Aggressive”.

**Conclusions**

As indicated in the previous discussion – and looking back at the initial problem statement: “Wilde’s RHT: Accept, reject or modify?” the following concluding statements are suggested:
• Wilde’s RHT must be acknowledged as an important stimulant for discussions and for stating problems and research projects associated with risk compensation and behavioural adaptation in the road traffic system.

• One main obstacle for a wide acceptance is the strict maintenance of the homeostasis concept which is not considered as a fruitful position to uphold. In this respect, the much more adequate concepts of risk compensation and behavioural adaptation are suggested as alternatives of rephrasing the RHT.

• By focusing less on utility concepts and more on compensatory mechanisms involved in information processing and decision-making and how behaviour is adapted to the variety of conditions in the road traffic system, it is believed that research will benefit considerably from taking into account the achievements of neuroscience.

• Wilde’s proposition of a target is a unique contribution from the Risk Homeostasis Theory, but also this concept needs to be rephrased from target risk to target feeling which seems more appropriate regarding what drivers actually are seeking in road traffic.

• Finally, RHT’s insisting of a target level of risk > 0 should be abandoned and replaced by a target risk = 0 which definitely is more in line with what the majority of drivers actually accomplish: It is more normal to drive a whole lifetime without personal injury accidents than the opposite.

References:


