The influence of environmental enrichment on the statement "happy animals make good science"

Master Thesis

Sanne van Gastelen

Neuroscience and Cognition, track Behavioural Neuroscience
Graduate School of Life Sciences
Utrecht University

Supervisor: Bart B. Houx

Overview

Introduction

Laboratory animal science is a field of study with the goal to contribute to the quality of research and to the wellbeing of laboratory animals. This field of study focuses on six different topics: (1) the biological characteristics of laboratory animals (i.e. in order to choose the right animal model), (2) the husbandry and housing conditions, (3) the genetic and microbiological quality, (4) the prevention and treatment of diseases, (5) the experimental procedures, and (6) anaesthesia, analgesic and euthanasia of laboratory animals (van Zutphen, 2003). Through the knowledge of these six different topics, laboratory animal science contributes to the reduction of distress experienced by the laboratory animals involved, and to the reliability and reproducibility of the results from animal experiments (van Zutphen, 2003).

These animal experiments are only allowed when there are well founded reasons for it. This means that the number of animal experiments should be reduced as much as possible and, if performed, these animal experiments should be executed with the greatest care (de Greeve and de Leeuw, 2003). In the Netherlands, these regulations have been statutory in the *Wet op de Dierproeven* in 1977 (rewritten in 1996) and in the *Dierproevenbesluit* in 1985 (rewritten in 1996). Both these laws stimulate the development of laboratory animal science in the direction which has been suggested by Russell and Burch (1959). These authors argued that scientists who use laboratory animals should always strive to replacement, reduction and refinement (i.e. the three R's). With replacement, Russell and Burch (1959) refer to the replacement of laboratory animals with alternative research methods that lead to same result but without the use of animals.

Russell and Burch (1959) defined reduction as the reduction of the number of laboratory animals needed for an experiment. This reduction can be reached through the use of a good research plan, the standardization of the laboratory animals used (e.g. genetic composition), and the standardization of the experimental procedures (van Zutphen, 2003). The last concept refinement is defined as the reduction of distress experienced by the laboratory animals. This would include the idea to optimise the housing conditions, the use of anaesthesia and analgesic, and good animal care (Russell and Burch, 1959; van Zutphen, 2003).

These three R's (i.e. replacement, reduction and refinement) form the backbone in the search for alternatives for laboratory animal science. Unfortunately, not all animal experiments can be replaced by alternative research methods. Thus as long as animal experiments need to be conducted (e.g. the search for medicine and the control of sera and vaccines), it is the duty of every scientist to reduce the number of laboratory animals needed and to refine the distress experienced by the laboratory animals involved.

In 1997 Poole argued that scientists are obligated to do everything practicable to ensure the happiness of laboratory animals, because the reduction of distress experienced by the laboratory animals does not only result in a better animal welfare, but also in good science. Poole (1997) stated that 'happy' animals (i.e. healthy animals whose welfare is not compromised, with normal behaviour and physiology) make good science. In the same article, Poole mentioned that "a factor, which is increasingly being recognized as a source of unhappiness, is the failure of the captive environment to meet the animal's behavioural needs and assure its psychological wellbeing. It is becoming apparent that captive animals can be bored or resort to abnormal behaviour if their environment is not sufficiently complex and interesting to them."

A possible solution to the fact that housing condition can be a source of unhappiness (i.e. animals whose welfare is compromised), is the application of environmental enrichment. Environmental enrichment can be defined as any modification or change of the environment in order to enhance the physical and psychological wellbeing of the animals involved, by providing stimuli meeting the species-specific needs of these animals (Baumans, 2000, 2005; Newberry, 1995). The provision of environmental enrichment potentially results in 'happy' animals. These 'happy' animals tend to have a normal physiology and express normal behaviour, and therefore, these animals are considered to be good subjects for scientific research (Poole, 1997).

‘Unhappy’ animals, on the other hand, live in distressing conditions (i.e. standard laboratory cages with no environmental enrichment), and tend to have behavioural and physiological disabilities, making them unsuitable subjects for scientific studies. These findings suggest that scientists should provide environmental enrichment to ensure the happiness of laboratory animals if the quality of their research is to be beyond reproach (Poole, 1997). But is it evidence based that ‘happy’ laboratory animals are suitable subjects, that ‘unhappy’ animals are unsuitable subjects, and that the provision of environmental enrichment indeed results in ‘happy’ animals?

Furthermore, concern has been raised that introducing environmental enrichment into the standardized cages of laboratory animals may increase the variability, resulting in an increase in the number of animals needed in order to achieve appropriate statistical power and an increase in the difficulty of duplicating the study in another laboratory (Bayne, 2005; Tsai *et al.,* 2002; Richter *et al*., 2009). Normally, to guarantee reproducibility of experimental outcomes, it is generally advised to standardize the conditions (i.e. no environmental enrichment) of the experiments. Environmental standardization reduces variation in the data, thereby increasing test sensitivity. Because higher test sensitivity allows a reduction of sample size, standardization is promoted for ethical reasons also in view of reducing animal use (Richter *et al*., 2009). Although it has been agreed that environmental enrichment has a beneficial effect on animal welfare (i.e. introducing refinement), it remains questionable whether the provision of environmental enrichment results in good science.

Ultimately, the decision to include environmental enrichment should be based on the consideration whether the enrichment has a beneficial effect on the animal, and whether the potential effects of the enrichment are experimentally relevant (Bayne, 2005). Therefore, this thesis will discuss several questions in order to evaluate the influence of environmental enrichment on the statement of Poole (1997), “happy animals make good science”. First, attention will be paid to the concept animal welfare and the possible influence of the housing condition on the welfare of laboratory animals. This will be done in order to answer the questions whether housing conditions can influence animal welfare and whether standard housing conditions can result in unhappiness. Then, the concept environmental enrichment will be discussed in order to answer the question whether the application of environmental enrichment in standard laboratory cages results in ‘happy’ animals with a better animal welfare. Before discussing whether environmental enrichment results in good science, with regard to its potential influence on the variation of the experimental results and the reproducibility of the experiment, the statement “happy animals make good science” will be evaluated. Is this statement evidence based? And is Poole’s theory accepted by other scientists or not, and do they provide evidence or examples to support or weaken this theory? Thus, all together, this thesis will discuss the influence of environmental enrichment on “happy animals make good science”.

Animal welfare and housing conditions

In his article, Poole (1997) states that "a factor, which is increasingly being recognized as a source of unhappiness, is the failure of the captive environment to meet the animal's behavioural needs and assure its psychological wellbeing. It is becoming apparent that captive animals can be bored or resort to abnormal behaviour if their environment is not sufficiently complex and interesting to them". He refers to unhappy animals as animals that live in distressing conditions and tend to have behavioural and physiological disabilities. In order to evaluate these statements, more information is needed about the concept animal welfare. What is animal welfare? Can housing conditions compromise animal welfare? And, what happens with the animals whose welfare is being compromised?

Animal welfare

Animal welfare is the physical and physiological wellbeing of animals, based upon the belief that animals are sentient (i.e. the ability to feel and perceive; Balcombe, 2010), and hence, able to experience pain and pleasure (Poole, 1997). Often, animal welfare has been characterized as a state of mental and physical health, indicating that the animal is living in harmony with its environment (Wiepkema and Koolhaas, 1993). In line with this characterization, Broom (1996) defined animal welfare as followed: "the animal's state as regards to its attempts to cope with its environment. This state includes how much the animal has to do to cope, the extent to which it is succeeding or failing to cope, and its associated feelings".

The Brambell Committee (1965) proposed that all animals should benefit from minimal standards of welfare, known as 'The Five Freedoms':

1. Freedom from hunger, thirst and malnutrition
2. Freedom from physical and thermal discomfort
3. Freedom from pain, injury and disease
4. Freedom from fear and distress
5. Freedom to express most normal patterns of behaviour.

These five freedoms mean that in order to benefit from minimal standards of animal welfare, an animal should be free from certain negative stimuli (i.e. the first four freedoms) and able to perform species-specific behaviours (i.e. the last freedom). Species-specific behaviours can be defined as behavioural patterns that are inborn in a species and performed by all members under the same conditions. These behavioural patterns are not modified by learning and all members of the species are highly motivated to perform these behavioural patterns (Lawrence, 2005).

In addition to these five freedoms, more scientists have given their idea about when animal welfare is present. It is, for instance, often believed that animal welfare is present when the animal is able to adapt to and cope with its environment, thereby reaching an internal state the animal perceives as positive. Another idea is that animal welfare is present when an animal can reliably predict and control relevant events by means of species-specific behaviours (Wiepkema and Koolhaas, 1993). This last theory will be described in more detail in order to provide more information about animal welfare.

It has been argued that stress responses are evoked when an animal is not able to foresee or control negative or positive events (Wiepkema and Koolhaas, 1993). This claim includes two important concepts: (1) predictability and (2) controllability. It may be helpful to explain these two concepts shortly. Some events can be predicted, but never actively controlled (e.g. arrival of a predator). Animals can learn to predict these types of events, when these events are preceded by reliable signals. Due to these reliable signals and hence predictability, animals may bring some passive control over these events through the performance of anticipatory actions (e.g. seeking shelter, hiding). Other events, however, can be fully controlled since their occurrence depends on the actions of the animal itself (e.g. availability of food) (Wiepkema and Koolhaas, 1993).

Thus, changes in the predictability and controllability can evoke stress responses. In a period during which the predictability and controllability are low, it can be assumed that the stress responses are severe and welfare is absent. On the other hand, when the predictability and controllability are high, the certainty of events is maximized and the stress responses are low. However, this does not necessarily mean that the animal welfare is good, because the high predictability and controllability also implies the near absence of novelties in the environment. Such a non-changing environment does not provide new and interesting information an animal needs and gathers during its exploratory behaviour. This will introduce boredom associated with a decreased vigilance (Wiepkema and Koolhaas, 1993). Thus, in other words, the predictability and controllability should not be too high or too low for long periods of time; it should have an intermediate value.

This intermediate level of predictability and controllability implies novelty and uncertainty and may evoke stress responses. However, the quality and quantity of these novelties and uncertainties are normally within the range of coping capabilities of the animals, because they have been adapted to environmental conditions which are not entirely stable (Wiepkema and Koolhaas, 1993). Thus, a baseline occurrence of behavioural and physiological stress responses is normal and does not necessarily reflect adverse or unacceptable conditions (Wiepkema and Koolhaas, 1993). These unpredictable and uncontrollable events may even be necessary to maintain normal vigilance.

Stress responses include a behavioural and a physiological aspect. When an unpredictable and uncontrollable event occurs, animals may express so-called conflict behaviours. These conflict behaviours include agonistic behaviour (i.e. a mixture of aggressive, threat and flight behaviour), displacement or interruptive behaviour, redirected behaviour and intention movements (Wiepkema and Koolhaas, 1993). However, an environmental challenge also induces a physiological response. This physiological response is initiated and coordinated by the central nervous system, which has two pathways; (1) the autonomic nervous system, and (2) the neuroendocrine system. One major subdivision of the autonomic nervous system is the sympathetic branch, which is activated by a stressor and makes the animal ready for physical activity (i.e. also called the fight-flight reaction) (Wiepkema and Koolhaas, 1993). This activity of the sympathetic branch, results in a wide range of physiological responses: an increased heart rate and blood pressure, an increase in the plasma (nor) adrenaline levels and an elevation of the body temperature. At the same time, a stressor may activate the neuroendocrine system, which results in a higher plasma level of corticosteroids.

Up to this point, the physiological and behavioural responses to an environmental stressor have been presented as highly functional, enabling an animal to cope with environmental challenges. In other words, these responses reflect the processes the animal uses to control its environment (i.e. to reach and maintain homeostasis). However, what happens when the changes in predictability and controllability become more drastic and permanent?

Signs of impaired welfare

When there is a short change in predictability and controllability, it will result in a strong activation of physiological and behavioural responses, but only for a short period of time. However, when the controllability and predictability changes permanently, the environment becomes beyond the control of the animals; they cannot adapt to and cope adequately with their environment. At this point, a long-term mild activation of behavioural and physiological responses has taken place, leading to chronic stress and welfare problems (Wiepkema and Koolhaas, 1993). When this is the case, a pathologic state will be reached which is characterized by physiological responses and behaviours which have no adaptive value. Thus, a variety of symptoms will develop, indicating chronic stress. Behaviourally, conflict behaviour gradually changes into abnormal behaviour (e.g. redirected behaviour may change into a stereotypy), and physiologically, the baseline level of physiological measures becomes chronically elevated (e.g. chronically enhanced plasma corticosteroid levels and abnormal reactivity of the neuroendocrine systems) (Wiepkema and Koolhaas, 1993).

From this point on, the main focus of this thesis will be on the behavioural responses of animals. Therefore, more information will be provided about the abnormal behaviours which develop when animals can no longer cope with their environment and when their welfare is compromised. Signs of impaired welfare are the suppression of normal behavioural patterns, abnormal behaviours, and fear and stress-related responses. The suppression of normal behavioural patterns can occur in the form of inactivity (i.e. suppressed activity). Abnormal behaviours can be divided into two categories: maladaptive and malfunctional behaviours (Mills, 2003). Maladaptive behaviours are performed by a normal animal in an abnormal environment. This means that an animal is responding as well as it can with functionally intact behaviours to the captive environment, and thus the animals' behaviour reflects the abnormality of the environment (Garner, 2005). Malfunctional behaviours, on the other hand, reflect the abnormal psychology, brain development, or neurochemistry of the animals, which are induced by features of the captive environment (Garner, 2005).

Abnormal repetitive behaviours are inappropriate, repetitive behaviours, unvarying in either goal or motor pattern (Garner, 2005). These abnormal repetitive behaviours can be divided into two categories based on the unvarying manner in which the behaviours are repeated. The first category is stereotypies, which are an unvarying inappropriate repetition of a particular set of movements and/or body postures that lack any goal or function (Garner, 2005). The second category is impulsive/compulsive behaviours, which are variable flexible goal-directed behaviours with the repetition of an inappropriate goal (Garner, 2005).

In mice, stereotypies include barmouthing (i.e. the mouse makes a series of repetitive mouthing movements on a cage bar) and jumping (i.e. the mouse rears on its hind legs and repeatedly jumps vertically on its hind legs, usually balancing on its tail, which is held rigid) (Würbel and Garner, 2007; Garner, 2005; Garner and Mason 2002). Both stereotypies are shown in figure 1a. An example of an impulsive/compulsive behaviour seen in mice is barbering (i.e. one mouse plucks similar patterns of fur from its cagemates or from itself if non-socially housed) (Würbel and Garner, 2007; Garner, 2005; Garner *et al*., 2004). Barbering is shown in figure 1b.

A B

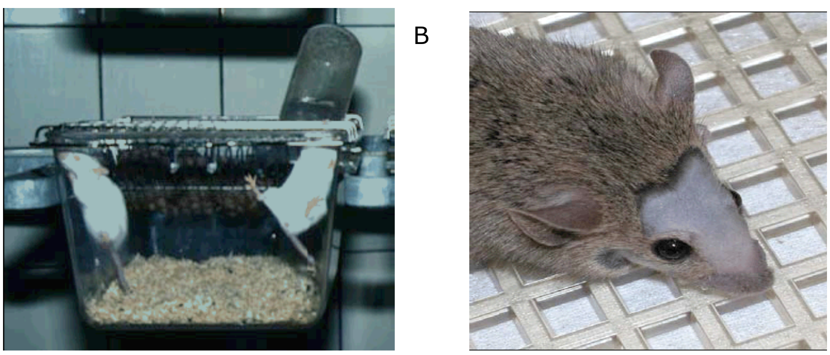


Figure 1. From Würbel and Garner, 2007. (a) Stereotypic jumping (let side) and bar-mouthing (right) in mice. Photo by H. Würbel. (b) A mouse that has been barbered by a cage mate. Photo by J.P. Garner.

Mason and colleagues (2007) proposed that stereotypic behaviours can be classified in two categories. The first category is induced by frustration and repeated attempts to cope, and can be called frustration-induced stereotypic behaviour. These frustration-induced stereotypic behaviours are driven by motivational frustration, fear or physical discomfort and reflect the nature of the underlying problem. These behaviours are not the product of any underlying dysfunction, but are deriving from attempts to replace a missing normal behaviour, to escape from confinement, or to otherwise alleviate the problem (Mason *et al*., 2007). The second category is induced by brain dysfunction, and can be called malfunction-induced stereotypic behaviour. These malfunction-induced stereotypic behaviours are products of a brain abnormality, and the behavioural patterns may be more arbitrary/less naturalistic, not directly reflecting the primary cause of the problem (Mason *et al*., 2007).

In summary, when animals live in an environment which is beyond their control, their welfare is compromised and abnormal behaviours are induced. As a result, abnormal behaviours are believed to reflect animal welfare (Mason, 1991), and therefore, often used to assess animal welfare. However, do abnormal behaviours always indicate a compromised animal welfare?

Are abnormal behaviours a sign of impaired welfare?

There is evidence that stereotypies indicate poor welfare, however, there is also evidence that stereotypies do not always indicate poor welfare (Mason and Latham 2004). For instance, some stereotypies increase in response to changes that can be presumed to be positive for animal welfare (e.g. Korhonen *et al.*, 2001), and stereotypies do not always positively correlate with other signs of poor welfare (e.g. corticosteroid level; Vestergaard *et al.*, 1997).

Mason and Latham (2004) reviewed almost 300 papers in order to see if stereotypies could be linked with poor welfare. They found that certain aversive environments enhance stereotypic behaviour, while at the same time, stereotypies can appear or increase in conditions that seems neutral or even beneficial (with respect to animal welfare) or that some aversive environments did not elicit stereotypies at all. Thus, it seems that the relationship between stereotypies and animal welfare is not straightforward (Mason and Latham, 2004). This might be explained with four theories.

The first theory is that stereotypies can act as 'do-it-yourself enrichments', by which the performance of stereotypies is linked with apparent benefit and the stereotypies have the ability to surrogate for natural behaviour. Thus performing these stereotypies can help laboratory animals to improve their own welfare by minimising the potential adverse effects of the environment (Mason and Latham, 2004). The second theory is that stereotypies can have a mantra effect, meaning that the performance of the stereotypies may serve as a coping mechanism, because of its positive reinforcing properties (Mason and Latham, 2004).

The third theory is that stereotypies can become habits due to the role of central control. Repetitive behaviours (e.g. stereotypies) may shift into a form of automatic processing, which is known as central control. This allows animals to execute regularly performed or fast movements with minimal cognitive processing or the need for sensory feedback. Such behavioural patterns may therefore become more easily triggered by a range of cues. Stereotypies that have reached this stage can be performed in a more diverse set of conditions and will be harder to interrupt or modulate with changes in the environment (Mason and Latham, 2004). This could result in an increased bout length and thence overall stereotypy levels, without any change in welfare.

The last theory, influencing the relation between animal welfare and abnormal behaviours, is that some animals can have a brain dysfunction that impairs the proper regulation of behaviour, resulting in animals producing behavioural responses to environmental cues that may be unnecessary or inappropriate.

The authors concluded that stereotypies undoubtedly have a role in welfare assessment. Thus stereotypies should always be taken seriously as a warning sign of compromised animal welfare, but never used as a sole index (i.e. thus in combination with other welfare indicators such as physiological measures). Furthermore, scientists should never overlook or assume that non-stereotyping or low-stereotyping animals are doing well (Mason and Latham, 2004).

So far, it has been discussed that the welfare of animals is compromised when the environment is beyond their control due to the permanent and drastic changes in controllability and predictability, and that the occurrence of abnormal behaviours should be taken seriously as a sign of impaired welfare. But when is such an environment reached in which the welfare of animals is compromised?

Housing conditions

An important finding is that abnormal behaviours do not develop in wild animals, neither in wild-caught adults (Callard *et al.*, 2000). This indicates that abnormal behaviours only develop in captive animals (i.e. zoo, farm and laboratory animals) and that an early critical period in captivity is required for the developed of these abnormal behaviours (Callard *et al*., 2000). Due to the fact that this thesis is discussing the statement of Poole (1997) (i.e. ‘happy’ animals make good science), from this point on, attention will only be paid to laboratory animals.

There is evidence that laboratory animals do not only suffer from experimental procedures, but also from their day-to-day living conditions (Balcombe, 2010). Ideally, laboratory animals should be housed in a complex, challenging environment that they can control (Poole, 1998). However, housing systems for laboratory animals have been designed according to hygienic, economic and ergonomic requirements to provide standardized conditions with little or no consideration for animal welfare (Baumans, 2004, 2005; Olsson *et al*., 2003). Due to a nutritionally well-balanced diet, a controlled climate and good hygienic conditions, the laboratory animals are kept in good physical health (Olsson and Dahlborn, 2002). However, their behaviour is severely restricted under standard housing conditions, and therefore, it is not surprising to find signs of impaired welfare in laboratory animals.

This type of housing condition (i.e. small, standard, barren cages) provides limited stimuli, prevent laboratory animals from exerting control over their situation, and gives limited opportunities for the animals to perform species-specific behaviours (e.g. social behaviour, exploring, seeking shelter, foraging, gnawing and nest building) (Baumans, 2005; Olssen and Dalhborn, 2002; Balcombe, 2006; Jennings *et al*., 1998; Latham and Mason, 2004). These species-specific behaviours can be considered behavioural needs; necessary for the maintenance of a normal physiological and psychological state (Poole 1998). Animals will be highly motivated to perform these species-specific behaviours and if the animals are not able to perform these highly motivated behaviours, their welfare will be compromised (Dawkins, 1998). Furthermore, laboratory animals may suffer from boredom in this restricted environment, because there are only a few opportunities to acquire information or interact with the environment (Poole, 1998). Both the thwarting of important behaviours as well as the lack of stimuli may be problems for animal welfare (Olssen and Dahlborn, 2002; Balcombe, 2010).

When laboratory animals are housed in these standard, barren environments, lacking stimuli an unable to perform species-specific behaviours, which would normally allow them to control their environment and enhance homeostasis, the animals will likely to be under a state of stress (Olsson and Dahlborn 2002). Restrictive housing conditions can cause stress in several ways (Olssen and Dahlborn, 2002): (1) the behavioural restriction itself may act as a stressor when highly motivated behaviours are thwarted, and (2) restrictive conditions limit the animals’ potential for controlling their physical and social environment (Wiepkema & Koolhaas, 1993). As previously mentioned, the inability to predict or control stimuli can lead to chronic states of stress (Garner, 2005).

Thus, when laboratory animals are housed in an environment where they are chronically exposed to aversive stimuli, or where they are unable to perform species-specific behaviours that would correct a homeostatic imbalance they are experiencing, it would result in abnormal behaviours and chronically elevated baseline levels of physiological measures (Wiepkema and Koolhaas, 1993; Garner, 2005; Olssen and Dalhborn, 2002; Würbel and Garner, 2007).

Figure 2. From Balcome, 2010; picture taken by C. Sherwin. The traditional standard, barren laboratory cages, experienced from the laboratory animals’ point of view (i.e. mouse).

Concluding remark

Animals continuously try to adapt to environmental conditions using behavioural and physiological mechanisms. Laboratory animals are subjected to the conditions offered by their caretakers. Under these controlled conditions, the originally highly adaptive physiological and behavioural mechanisms may be no longer functional, the opportunity of performing species-specific behaviours is limited and there is a lack of stimuli. This together can lead to a decrease in welfare and finally to pathology. Thus, to come back at the beginning of this chapter, Poole (1997) was indeed right that standard laboratory housing conditions can result in compromised welfare, leading to animals with behavioural and physiological disabilities.

However, a new question arises. Poole (1997) argues that animals whose welfare is being compromised are unsuitable subjects for science due to their abnormal physiology and behaviour, and that conclusions based on these animals might be unreliable. According to Poole (1997) only animals whose welfare is not compromised are suitable subjects and make good science. However, as just mentioned, standard laboratory cages unfortunately often result in animals whose welfare is compromised. Thus, the happiness of laboratory animals seems to be restricted by their housing conditions. Is there perhaps a solution to reduce or prevent the abnormal behaviour and physiology, and hence increasing the happiness of these animals?

Environmental enrichment

As mentioned in the last paragraph of the previous chapter, there seems to be a problem between the standard laboratory housing conditions and the happiness of laboratory animals. Animals whose welfare is not compromised are suitable subjects and make good science (Poole, 1997). However, most laboratories use standard cages, which often result in animals whose welfare is compromised, due to the fact that these cages have been designed according to hygienic, economic and ergonomic requirements to provide standardized conditions with little or no consideration for animal welfare (Baumans, 2004, 2005; Olsson *et al*., 2003). Thus, a solution is needed to reduce and prevent abnormal physiology and behaviours, and increase welfare in laboratory animals in order to make good science based on healthy subjects with a normal physiology and behavioural pattern (Poole, 1997). Therefore, in this chapter a few possible solutions will be shortly discussed. Probably the most promising and most frequently used solution is environmental enrichment. Therefore, the focus of this chapter, and the rest of this thesis, will be on environmental enrichment.

Reduce or prevent signs of impaired welfare

Most often, attempts are made to reduce or prevent abnormal behaviours. There are five different types of attempts to do this: (1) genetic selection, (2) the use of pharmacological compounds, (3) positive reinforcement of alternative behaviours, (4) physical prevention or punishment, and, most commonly, (5) environmental enrichment (Mason *et al*., 2007). Genetic selection could be used in order to reduce abnormal behaviours through selecting against abnormal behaviour. Welfare could potentially be improved, by breeding in favour for animals that are not frustrated by captive conditions or vulnerable to stressors. However, there is a larger disadvantage, namely that selecting against abnormal behaviour itself rather than its underlying causes, could result in selecting for inactive phenotypes that may have an even poorer welfare (Mason *et al*., 2007).

The use of pharmacological compounds is the second method to reduce abnormal behaviours. Some have successfully reduced stereotypic behaviour, however, long-term pharmacological intervention seems a rather perverse solution to problems fundamentally caused by housing (Mason *et al*., 2007). The approach to reduce abnormal behaviours with actively reinforcing non-stereotypic behaviours with a reward, has been shown successful. However, similarly to pharmacological compounds, it may not tackle the underlying causes of abnormal behaviours (Bloomsmith *et al*., 2007). Furthermore, positive reinforcement of alternative behaviours can result in other behaviour patterns becoming repetitive, or placing an animal in a state of conflict (i.e. the motivations to perform an abnormal behaviour with the motivation to perform the rewarded behaviours) (Mason *et al*., 2007).

The effectiveness of physical prevention of abnormal behaviours is uncertain. This is due to the fact that animals simply shift their abnormal behaviours to a different location, incorporate the imposed obstacle into the behaviour (Vickery and Mason, 2003), or even increase their abnormal behaviours due to the resulting stress (Bloomsmith *et al*., 2007). Researchers warn for the potential dangers of blocking abnormal behaviours (e.g. Mason and Latham, 2004), given the risk that these approaches may prevent the behaviours that animals find stress-relieving or otherwise beneficial. And, also very important, this approach (i.e. physical prevention) clearly does not tackle the underlying causes of abnormal behaviours (Mason *et al*., 2007.

The last method, environmental enrichment, does tackle the underlying causes of abnormal behaviours (i.e. the housing conditions). Environmental enrichments are the modifications or interventions made to the environment with the aim of improving the welfare of the animals living in that environment. Environmental enrichment can reduce the expression of abnormal behaviour, increase the expression of normal behavioural patterns and increase the ability to cope with challenges in a more normal way (Rochlitz, 2005). In the next part of this chapter, the entire concept of environmental enrichment, in combination with its ability to improve animal welfare and its unintended consequences will be discussed.

The definition and goals of environmental enrichment

Environmental enrichment can be defined as any modification or change of the environment which increases the complexity of an enclosure and provided more stimuli in order to provide the opportunity for animals to perform species-specific behaviours, with the goal of improving animal welfare (Baumans, 2000, 2005; Young, 2003, Newberry, 1995; Hutchinson *et al*., 2005).

It is not correct to use the term environmental enrichment before the following results have been shown: (1) enhanced animal welfare, and (2) improved biological functioning of the animals. (Baumans, 2005). Thus, for environmental changes or modification to be considered enriching, its produced results (e.g. behavioural, psychological, physical and physiological) must be connected with an improvement of animal welfare. Therefore, it is important to make a distinction between environmental interventions that cause certain changes which may be considered to be welfare-neutral, and environmental enrichment for which evidence exist to suggest an improvement of welfare (Ellis, 2009). In order to qualify these environmental changes or modifications as enrichment, it has often been suggested (Baumans, 2005; Olsson and Dahlborn, 2002) that any change to the housing system should:

1. Improve the quality of the environment so that the animal has a greater choice of activity and some control over its social and spatial environment.
2. Increasing the frequency and diversity of natural, species-specific behaviours.
3. Reducing the frequency of abnormal behaviour.
4. Increasing the positive utilization of the environment.
5. Increasing the animal's ability to cope with challenges of captivity.

These quantifiable goals make it possible to evaluate the results of environmental enrichment strategies (Baumans, 2005; Ellis, 2009; Leach *et al*., 2000; Newberry, 1995; Olssen and Dahlborn, 2002; Stauffacher, 1995; Young 2003). In the following part of this chapter, a summary will be given about the different environmental enrichment strategies which are available for laboratory animals, and whether these strategies potentially improve animal welfare.

Different environmental enrichment strategies

Environmental enrichment strategies can be categorized into two types: (1) social and (2) physical enrichment (van de Weerd and Baumans, 1995; Young, 2003). Social enrichment can consist of direct interactions between animals through social housing, to mere visual contact or auditory communication between animals (Hutchinson *et al*., 2005). Thus, social enrichment includes socialization of animals both in contact and not in contact (i.e. noncontact) with conspecifics (Baumans, 2005).

With social contact enrichment scientists refer to housing animals in groups or in pairs with conspecifics. If this type of enrichment is used, it is important that the group composition is stable and harmonious (Morton *et al*., 1993; Stauffacher, 1997a; Turner *et al*., 1997), and that certain visual barriers or hiding places are provided in order to minimize aggression (Stauffacher, 1997b; Van de Weerd and Baumans, 1995; Van Loo *et al*., 2002). This is due to the fact that, even in harmonious groups, it is necessary to allow animals to initiate contact by approach or to avoid contact by withdrawal from sight (Baumans, 2005). With social noncontact enrichment, scientists refer to visual, auditory, and olfactory communication with conspecifics (e.g., through bars). The Council of Europe states, on the accommodation and care of laboratory animals, state that when group housing is not possible, "consideration should be given to accommodating conspecifics within sight, sound or smell of one another" (Council of Europe, 1997 and NRC, 1996, both referred in Baumans, 2005).

The other category of environmental enrichment is physical enrichment. This includes a complex environment with both sensory and nutritional stimuli. With complexity, scientists refer to an appropriate structuring of the environment in which animals use resources and structures in order to perform species-specific behaviours

(Baumans, 2005). The housing environment can be made more complex by the use of structures, substrates and manipulanda (Hutchinson *et al.*, 2005). Structures and substrates include any object or parameter of the environment that allows an animal to isolate itself into different microenvironments, experience varied textures or materials, or express natural patterns of locomotion. These enrichment strategies include nesting materials, running wheels, and hiding shelters (Hutchinson *et al.*, 2005). Manipulanda, on the other hand, include any object that can be altered by an animal or encourage it to engage in fine motor movements, such as wooden blocks or prefabricated plastic chew toys (Hutchinson *et al.*, 2005).

Physical enrichment can be achieved by providing sensory stimuli, including visual, auditory, olfactory, tactile, and taste stimuli. Probably, the most satisfying enrichment for rodents is visual, auditory, olfactory, and tactile communication with conspecifics, either directly or through bars (i.e. social enrichment) (Baumans, 2005). Although other forms of sensory enrichment are not commonly used, they do exist, such as the provision of acoustic and scent stimuli (i.e. auditory and olfactory) (Wells, 2009).

The last form of enrichment is nutritional enrichment. It is known that animals are highly motivated to make use of enrichment involving food items. This type of enrichment includes novel food items (e.g. fruits, vegetables and grains) which differ in taste, texture and desirability (Hutchinson *et al.*, 2005). Furthermore, the presentation of food, giving animals the opportunity to forage (e.g., scattering food in the bedding), is another form of nutritional enrichment. It appears that this last type of nutritional enrichment prevents boredom, and that nutritional enrichment in general satisfies three different types of species-specific behaviours: (1) exploration, (2) appetitive behaviours, and (3) mouth movements (Baumans, 2005; Mench, 1998).

As illustrated above, there are many different types of environmental strategies, all having the potential of improving the welfare of laboratory animals. However, before introducing environmental enrichment strategies into the laboratory cages, it is very important to determine which type of environmental enrichment is important for animals, and what the consequences would be for the animals. Do these environmental strategies improve animal welfare or do they have negative consequences?

Do environmental enrichment strategies improve animal welfare?

In order to determine which type of environmental enrichment is important for animals, two types of experiments can be conducted: (1) preference testing, and (2) motivational testing. In preference testing experiments, animals are given a choice between different conditions in order to examine the preference of the animals for specific objects and environments. The condition which is most frequently chosen is concluded to be the preferred strategy. This type of experiment is appealing, but it can only measure relative preferences between conditions and cannot establish the strength of the desire for the preference (Baumans, 2005; Benefiel *et al.*, 2005; Olssen and Dahlborn, 2002). Motivational testing addresses this problem, by measuring the amount of work an animal is willing to perform to gain access to a certain environment or object. This experiment assumes that animals will work harder to obtain enrichment conditions that fulfil their most basic needs (Baumans, 2005; Benefiel *et al.*, 2005; Olssen and Dahlborn, 2002).

However, it is important to note that preference and motivational tests assume that animals will always pursue their best welfare interests. This assumption may be wrong in certain cases, therefore, the preferences of animals and their motivation for certain preferences may not be the ideal guideline to what is of most value to their wellbeing (Benefiel *et al.*, 2005; Hutchinson *et al.*, 2005). Thus, next to investigating the preference of animals for which type of environmental enrichment, it is very important to examine the effects of these environmental enrichment strategies on behavioural or physiological parameters.

Olssen and Dahlborn (2002) reviewed 40 studies, performed between 1987 and 2000, to evaluate environmental modifications, and concluded that mice prefer a more complex environment containing nesting material, and that they were prepared to work for access to such preferred environments. The preference for nesting material might be due to that fact that it enables them to create appropriate microenvironments for resting. It has been shown that laboratory mice will use nesting material and perform nest-building behaviour (Eskola and Kaliste-Korhonen 1999; Van de Weerd et al. 1997a, 1998), and that they spent 10 to 20% of their time budget manipulating nesting material (Van de Weerd et al. 1997b).

Olsson and Dahlborn (2002) also concluded that housing mice under these preferred conditions (i.e. higher complexity and bedding material) improves animal welfare (Fraser and Matthews, 1997; Olssen and Dahlborn, 2002). Housing mice in a complex environment provides various opportunities for exploration and locomotion, resulting in behaviour changing significantly in the desirable direction (e.g. more play and positive social interactions, and fewer stereotypies and agonistic behaviour) (Nevison *et al.*, 1999). In addition, a higher complexity affects their reactions to many behavioural tests. Typical results were increased activity, decreased fearfulness, reduced signs of anxiety and a reduced latency to emerge (Olssen and Dahlborn, 2002; Chapillon *et al.*, 1999). Furthermore, mice housed in an environment with nesting material, consumed less food without a consequent decrease in their body weight (Dahlborn *et al.*, 1996, van de Weerd *et* al., 1997b; Olssen and Dahlborn, 2002). This effect is commonly attributed to the thermoregulatory properties of a nest. The presumed improved thermoregulation in mice with access to nesting material may be considered an improvement in biological functioning and is also expected to affect welfare positively (Olssen and Dahlborn, 2002).

Thus based on these reviewed studies, it can be concluded that both nesting material and higher complexity due to cage structures, have clear functions in that they provide shelter and give the opportunity to perform species-specific behaviours, such as nest building, exploration, climbing and vigilance. Additionally, controllability of the environment is increased, presumably resulting in less stressful conditions (Wiepkema and Koolhaas, 1993; Olssen and Dahlborn, 2002).

However, although intuition and even direct observation indicate that environmental enrichment promotes species-specific behaviours, data showing that rodents benefit from these changes are often inconclusive (Benefiel *et al.*, 2005; Olssen and Dahlborn, 2002). This means that the results (i.e. benefit for animal welfare) often vary between different species, different strains, different sex, different age, etc.

Concluding remark

This chapter may give rise to more question than providing answers. As mentioned previously, environmental enrichment is considered to imply an increase in the complexity of the environment in which the animal lives, with the goal of enhancing the animal’s welfare. Despite the clear example given above, it is often unclear whether environmental enrichment benefits an animal’s wellbeing, due to the fact that is not conclusive for all species or for all forms of environmental modifications/changes (e.g. strain, sex and age). In addition, the potential for environmental enrichment to have unexpected consequences such as unintended harm to the animal, or the introduction of variability into a study that may confound the experimental data, has received recent interest (Bayne, 2005).

Bayne (2005) illustrated, similar to Olssen and Dahlborn (2002), that environmental enrichment in some cases, has no impact on the animal or research parameter being studied, and in other cases environmental enrichment has an effect on either the animal or the research data (see following chapters) in either a positive or a negative manner, depending upon the specific circumstances. Furthermore, Bayne (2005) states that implementation of environmental enrichment should include an assessment of the potential risk to the animal associated with the item. All together, this indicates that the decision to include environmental enrichment should be based on a consideration of the safety of the animal, whether the enrichment has a demonstrable beneficial effect on the animal, and whether the potential effects of the enrichment are experimentally relevant (Bayne, 2005).

In summary, despite the lack of scientifically evidence and scientifically based guidelines, environmental enrichment is a widely accepted practice. In addition, many recent studies on rodent welfare support the view that environmental enrichment is by definition a good thing, without evidence that it is essential for animal welfare or consideration of its possible effects on experimental outcomes (Benefiel *et al*., 2005). Considering the effects environment enrichment can have on the physiology, brain structure and function, and the behaviour of animals, it makes sense that scientists have concerns about how these changes may affect experimental data and the validity of research results (Benefiel *et al*., 2005).

Environmental enrichment is an experimental variable and therefore, the use of enrichment may have effect on research data (Hutchinson *et al*., 2005). This have led to the concern that environmental enrichment conflicts with standardization of animal experiments because animals from an enriched environment show more variability in their response to experimental procedures, thereby potentially influencing the number of animals required and the reproducibility of the research. However, on the other hand, scientists who consider environmental enrichment as a good thing, tend to disagree. According to them, animals housed in an enriched environment tend to be less reactive to stressful experimental situations, hence resulting in less variation between results, which should reduce the number of animals used and enhancing reproducibility (Baumans, 1997; van de Weerd *et al*., 2002).

In conclusion, it is widely accepted that the appropriate application of environmental enrichment results in approved animal welfare, despite the lack of scientific evidence. Thus, coming back to the statement of Poole (1997), housing laboratory animals in enriched housing conditions (rather than in standard laboratory cages) will result in animals whose welfare is less compromised. This would, according to Poole (1997), mean that the animals housed in enriched conditions would be 'happy' and thus better subjects for science. However, due to the possible negative effects of environmental enrichment of the experimental data, this is not so straightforward. Thus, in the next chapter, it will be discussed whether animals whose welfare is compromised are unsuitable subjects and whether environmentally enriched animals make better subjects. And in the last two chapters, the potential negative effects of environmental enrichment on scientific outcomes (i.e. the potential greater variation among data and the potential effect of reproducibility) will be discussed.

Animal welfare and good science

So far, it has been shown that the housing conditions of laboratory animals can influence their welfare. Animals living in distressing conditions beyond their control, most often have behavioural and physiological disabilities (e.g. permanently increased levels of stress hormones and a compromised immune system), due to the fact that their welfare is compromised. Animals that live in enriched environments can have a better welfare resulting in a wide repertoire of behaviours, no/less display of fear towards trivial non-threatening stimuli and no expression of abnormal behaviour. Furthermore, a 'happy' animal is able to cope with stressors to which it is subjected. Up to this point, Poole (1997) was right. However, what about his statement "happy animals make good science"?

In order to evaluate this, it is important to describe what good science is and what determines the quality of experimental research. Without this description of good science, it would be difficult to investigate whether the use of 'unhappy' animals indeed results in poor science and whether the use of 'happy' animals indeed results in good science.

What makes good science?

In order to have a scientific research of high quality, the results of that research must be valid, reliable and replicable (Martin and Bateson, 2007). These three aspects of scientific research will be shortly discussed, in order to explain the possibility of 'unhappy' animals being unsuitable subjects. The first aspect, making a scientific research of high quality, is the reproducibility of the results. Reproducible results are results that can be repeated independently in different laboratories (Garner, 2005).

The validity of a result can be divided into two different types of validity: (1) internal and (2) external. A scientific result shows internal validity when the measured results measured what it was intended to measure, i.e. to how well a study was performed, how strictly confounding variables were controlled, and how sure a researcher can be that the found effects are caused by experimentally manipulating the independent variable(s), and not by confounds (van der Staay *et al*., 2009). Confounds are considered to be factors that potentially affect the independent variable as well and, may therefore, provide alternative explanations of the obtained results (Guala, 2003). In addition, a result shows external validity when the result obtained can be applied to other conditions (i.e. environmental contexts), populations or species (Campbell, 1957). Thus, external validity refers to the robustness of the results outside the circumstances in which it was established (Guala, 2003) and defines the extent to which a result can be generalized (Richter *et al*., 2009).

Reliability refers to the extent to which results are repeatable and consistent (i.e. free from random errors). To explain this in more detail, it is important to know that a result consists of two parts: (1) a systematic component, representing the true values of the variable, and (2) a random component arising from imperfections in the measurement process. When the random component is small, the results are more reliable. Thus, reliable results measure a variable precisely and consistent (Martin and Bateson, 2007). Two different aspects of reliability can be distinguished: (1) within-observer reliability (i.e. test-retest reliability) describing the extent to which the same experimenter obtains consistent results when measuring the same animals with repeated observations, and (2) between-observer reliability describing the extent to which two or more experimenters obtain similar results when measuring the same animal simultaneously (Garner, 2005; Martin and Bateson, 2007).

Thus, the desired outcome of every experiment is scientifically valid, reliable and reproducible data. With this clear description of good science, the following two questions need to be answered: (1) Will the use of 'unhappy' animals (i.e. whose welfare is compromised) prevent this desirable outcome? And (2), will the use of 'happy' animals (i.e. whose welfare is not compromised) result in this desirable outcome? According to Poole (1997), the answers to both of these questions are yes, resulting in his statement "happy animals make good science". In order to evaluate this, the evidence on which

Poole (1997) based his theory will be discussed, followed by the possible acceptance of this theory by other scientists.

"Happy animals make good science"

Poole (1997) tried to answer the question whether the state of mind of an animal would have the potential to influence scientific results which are derived from that animal. As mentioned previously, he distinguished two different states of mind: happy and unhappy. He also argues that in the modern laboratory science, many distressing factors have been eliminated (e.g. no illness, no injuries, and certain essential physical needs, such as food, water or suitable climatic conditions have been satisfied). However, a variety of potential causes of distress remain present, which can be categorized in three classes (1) social factors, (2) physical factors, and (3) handling and training. According to Poole (1997) these three potential distressing factors can influence the state of mind of an animal and, therefore, also the scientific results. This is due to the fact that scientific methods assume the absence of confounding factors, and clearly, 'unhappiness' can be a confounding variable. In the following paragraphs Poole's consideration about how these three potential distressing factors can influence the psychological wellbeing of the laboratory animals, will be illustrated.

When discussing the potential influence of social factors, Poole (1997) clearly distinguishes between mice and rats. While female mice can easily be socially housed in single sex groups, male mice tolerate these conditions less easily due to fact that they tend to fight and establish a hierarchy. However, the type of social structure depends entirely on the number of individuals present in the cage. Rats, on the other hand, are more sociable than mice and both female and male adult are likely to suffer from isolation. Overall, it is important to house animals in conditions where their social grouping leads to the minimum of aggression and hence distress. It has, for instance, been shown that subordinates male mice express higher levels of stress than dominants (Hucklebridge *et al.* 1976, Benton *et al.* 1978), that they experience more fear and that their immunological response was clearly reduced in comparison to that of dominants (Beden and Brain, 1984, 1985). Thus, even though the social companion ship may benefit mice in general, it is important to provide the opportunity for individuals to avoid another individual and minimize the chance for conflict, in order to prevent the higher levels of stress and fear, and a reduced immune response (Poole, 1997).

Furthermore, the environment in which laboratory animals live during their development determines the kind of situations with which they are able to cope when they reach adulthood (Poole, 1997). During their development, animals play fight in order to practise the strategies of attack and defence which they will need when faced with rivals. In this developmental phase, animals show curiosity and inventiveness, and thus learn the properties of objects and other organisms in their environment. Furthermore, animals enjoy play and experimentation, because it is self-rewarding (Poole, 1997). Thus, the provision of a stimulating and complex developmental environment will stimulate these developments of laboratory animals and enlarge the range of situations with which these animals can cope. If such an environment is not provided, it might result in abnormal development and a larger range of situation with which the animal cannot cope and thus potentially result in compromised welfare (Poole, 1997).

Likewise, the presence of a mother is important because it allows the young to express their wide repertoire of play and curiosity without fear. Therefore, early weaning is most probably stressful for animals (Poole, 1997). The problem becomes larger with the sudden loss of maternal antibody. This immune deficiency is temporary until the young independent develops its own fully functioning immune system. However, the temporary immunodeficiency may be enhanced by stress resulting from separation from their mother (Poole, 1997). These effects of early weaning can potentially result in animals developing abnormal behaviours and potentially suffering from reduced immunological competence as adults.

The second potential distressing factor influencing the state of mind of an animal is the physical environment in which the animal is housed. As already mentioned in the chapter welfare, Poole (1997) states that differences in the level of stress hormones is clearly related to the controllability and predictability of events. In addition, Wiepkema and Koolhaas (1993) argued that an intermediate level of predictability and controllability in needed to maintain a certain level of vigilance. Thus, a baseline occurrence of behavioural and physiological stress responses is normal or natural and does not necessarily reflect adverse or unacceptable conditions. However, it is important to realise that certain stressors can seriously compromise experiments (e.g. poor husbandry, such as continuous illumination and loud noises, can influence the appetite of mice and rats for sweet substances and thus compromise experiments using a food reward; Monleon *et al.*, 1995).

Furthermore, the environment contains stressors of which we are unaware, that can potentially influence the physiology, immune system and behaviour. Some of the most stressful events in every day husbandry result from changes in environment. Animals may be placed in unfamiliar/clean cages and, for animals which mark their home range with scent, this may be highly disturbing. Even more stressful is the situation where an animal is moved from its home cage to an unfamiliar one and then subjected to an experiment. Therefore, it is advised to provide time for the animal to acclimatize to its new situation, not only on welfare grounds but also because the experimental results may be influenced (Poole, 1997). This advice is supported with an experiment from Damon and colleagues (1986), which compared the nephrotoxic response of rats to implanted, refined uranium ore. They found that for rats who were not acclimatized 3-8mg/kg proved toxic, while for the acclimatized rats, the toxic dose was 220-650 mg/kg. This data support the practice of carrying out experiment in a familiar environment.

The last potential distressing factor influencing the state of mind of an animal is the way animals are handled. It is, for instance, well known that restraint can be highly stressful for animals resulting in increased levels of corticosteroids, and it may also lead to the suppression of their immune response (Lawrence, 1991; Blecha *et al.*, 1982). In addition, laboratory mammals recognize humans as individuals and are nervous of strangers. Therefore, it would be preferable that an experimental procedure is carried out by a handler which is a person familiar to the animal, and in whom the animal has confidence (Poole, 1997). An unfamiliar handler will most likely cause the animal to experience fear and stress.

In addition, when laboratory animals are good handled and trained to cooperate, it will improve the quality of the relationship between carer and animal, the animal will gain confidence and trust in the handler, and the animal will be less stressed. This all together, will result in improvement of the experiment, because the unwanted variable (i.e. stress) has been reduced. Thus a positive, caring attitude by staff not only improves the wellbeing of the animal but also makes it more willing to cooperate in any procedures to which it is subjected.

In summary, Poole (1997) gives examples which show that both the endocrine condition and immunology of laboratory animals, which experimenters may assume to be normal, can be compromised by social conditions, developmental history and stressors in the animal's housing condition or experimental laboratory. On these examples, he based his conclusion that happy animals are healthy and have a normal physiology and behaviour. These animals are able to cope with the stressors to which they are subjected, and will not be in pain, or distressed and resist disease. If these happy animals receive a good and simple treatment, stress will be even more reduced and animals become confident, cooperative and easily handled. All this together, makes these happy animals the best subjects for scientific research (Poole, 1997). In addition, he concluded that unhappy animals have to put up with distressing conditions beyond their control, resulting in behavioural and physiological disabilities (e.g. permanently raised levels of stress hormones and a compromised immune system). These uncontrolled variables make unhappy animals unsuitable subjects for scientific studies (Poole, 1997).

Based on the examples Poole (1997) gives, his conclusions seem logical and valid. He has shown that certain factors (e.g. social, physical and handling) can influence the state of mind of animals. If animals are not housed in the right social condition, if animals have to deal with uncontrollable, unpredictable husbandry events, and if animals are not treated kindly and gently, the welfare of these animals can be compromised resulting in 'unhappy' animals with behavioural and physiological disabilities. Unhappiness is a confounding factor because it consists of uncontrolled variables, and therefore, using these 'unhappy' animals will most probably not result in scientifically valid, reliable and reproducible data. This indicates that using 'unhappy' animals result in poor science and the use of 'happy' animals result in good science.

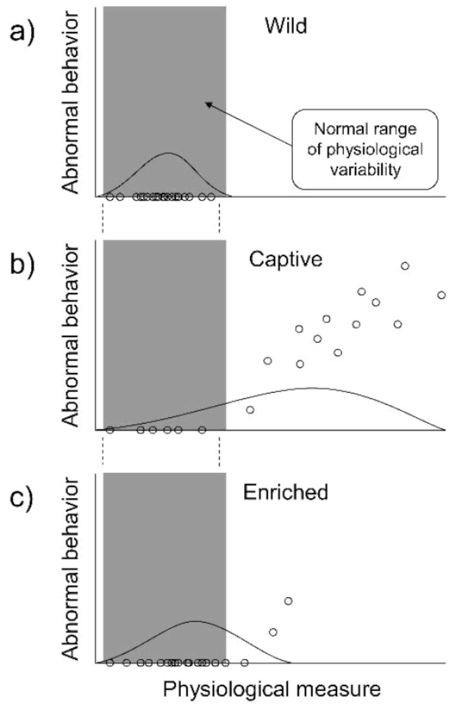
It would be interesting to illustrate how other scientists think about this theory thirteen years after it has been published by Poole (1997). Is Poole's theory accepted by other scientists or not, and do they provide evidence or examples to support or weaken this theory?

Acceptance "Happy animals make good science"

If laboratory animals are housed in an environment in which they are exposed to chronic aversive stimuli, in which the performance of species-specific behaviours is limited, or in which the performance of behaviours that allow the animal to control its environment and enhance homeostasis is limited, they are likely under a state of stress, resulting in the development of abnormal behaviours (Wiepkema and Koolhaas, 1993; Baumans, 2005; Olsson and Dahlborn, 2002). Thus, as Poole already stated; "the failure of the captive environment to meet the animal's behavioural needs and assure its psychological wellbeing, is an important source for unhappiness. It is becoming apparent that captive animals can be bored or resort to abnormal behaviour if their environment is not sufficiently complex and interesting to them".

Garner (2005) argued that animals performing abnormal behaviours could result in the reduction of reliability, reproducibility and the validity of scientific results, thereby supporting the theory of Poole (1997). According to Garner (2005), there are two important aspects of abnormal behaviours: (1) prevalence and (2) symptom severity. The first aspect, prevalence, describes the fact that some individuals perform abnormal behaviours, while others do not, even when they are of the same strain, sex, and age, experiencing the same housing, husbandry, and handling, and housed in the same cage. The second aspect, symptom severity, describes the fact that individuals that perform abnormal behaviours, differ significantly in the severity of these behaviours. These differences in abnormal behaviour between individuals, sites and strains, indicate that abnormal behaviours represent an uncontrolled variable which can potentially add considerable between-individual noise to experimental results, resulting in reduction of the reliability of experimental results (Garner and Mason, 2002).

Including animals with physiological and behaviour disabilities into the experiment can have an affect on three aspects of an experiment, validity, reliability and reproducibility. The validity can be affected due to the introduction of abnormal animals into the experiments, the reliability can be affected due to the increased inter-individual variation through the introduction of abnormal animals and the reproducibility van be affected by altering the number and type of abnormal animals between laboratories (Garner, 2005). This idea is presented in figure 3. The grey box in this figure, describes the range of physiological variation within a wild, 'normal' population that can be considered normal. This range of physiological variation should be the target range for a valid study. The first graph (i.e. figure 3a) shows the distribution of a physiological measure in a wild, 'normal' population of animals. None of these animals show abnormal behaviour. Figure 3b presents the distribution of a physiological measure in a captive population. Within this captive population, the environment has induced abnormal behaviours and an abnormal physiology, resulting in only a few individuals within the normal range of physiological variation.



Thus, only a small proportion of animals whose welfare is compromised, have a physiological measure within the target range. Using these animals could lead to poor validity, because the abnormal behaviours and physiology are considered to be factors that potentially affect the independent variable and, may therefore, provide alternative explanations of the obtained results (Guala, 2003). Furthermore, the increase in physiological variation will decrease the reliability of any experiment performed with animals whose welfare is compromised, because a reliable result measures a variable precisely and consistent and is free from random errors. If the animals perform abnormal behaviour and have an abnormal physiology, the imperfections in the measurement process increases, and hence, the random component increases. Thus, the results become less reliable (Garner, 2005).

Figure 3. From Garner, 2005. A hypothetical data set illustrating the idea that animals performing abnormal behaviour can potentially influence the validity, reliability and the reproducibility of scientific results. (a) represents a wild, 'normal' population, (b) represents a captive population, and (c) represents a captive population in an enriched environment.

Due to the variation in the severity and the prevalence of abnormal behaviour, as discusses beforehand, scientists should either quantify abnormal behaviours, using them as an explanatory variable (i.e. similar to sex, age and weight), or, preferably, modify housing conditions to prevent abnormal behaviours (Garner *et al*., 2003). As discussed in the chapter environmental enrichment, animal welfare can potentially be improved and the occurrence of abnormal behaviours decreased with the application of environmental enrichment (Würbel, 1998; Olsson and Dahlborn, 2002). Therefore, environmental enrichment might result in animals being 'happy' (i.e. no compromised welfare) and can potentially enhance the validity and reproducibility of animal experiments by normalizing abnormal physiology (Poole, 1997; Garner and Mason, 2002).

This normalization of abnormal physiology is shown in figure 3c. Animals housed in an enriched environment show less abnormal behaviours and, therefore, have a less abnormal physiology. Thus, due to environmental enrichment, the range of physiological variation becomes close to the range of physiological variation of a wild, 'normal' population, which, as previously mentioned, should be the target range for a valid study. Thus, by reducing the range of physiological variation, environmental enrichment could potentially reduce between-individual and between-laboratory variability, resulting in an increased experimental reliability and reproducibility.

It is important to note that this theory from Garner (2005) is entirely illustrated by a hypothetical data set. This means that there is no real evidence to support his illustrated theory. However, his theory is based upon previous research. It is known that abnormal behaviours and physiology does not occur in animals from a wild population (Callard *et al*., 2000). Thus is would be logical to assume that their range of physiological measures and behaviours can be assumed to be normal and considered a target for good science. As discussed in the chapter welfare, animals kept in captivity (e.g. standard laboratory cages) can have a compromised welfare, because of the restriction of species-specific behaviours, or the restrictive conditions limit the animals' potential for controlling their physical and social environment (Wiepkema & Koolhaas, 1993). Compromised welfare can result in abnormal behaviours and physiological disabilities, outside the range which was assumed to be normal (i.e. from wild population). As Poole (1997) already stated, behavioural and physiological disabilities are confounding factors because they consists of uncontrolled variables, potentially influencing the validity, reliability and the reproducibility of scientific results.

Furthermore, as discussed in chapter environmental enrichment, some environmental enrichment strategies (i.e. bedding material and higher complexity) can indeed result in improved welfare and thus reduce or prevent the occurrence of abnormal behaviour and physiology (Olssen and Dahlborn, 2002). It seems logical that the range of physiological measures of enriched animals become closer to the range of the wild population, which was considered a target for good science. Thus, using animals whose welfare is not / less compromised (e.g. through the application of environmental enrichment) could result in a higher validity, reliability and reproducibility of scientific research.

So far, there seems to be evidence that 'happy' animals whose welfare is not compromised are better subjects for scientific research than 'unhappy' animals whose welfare is compromised. Does this mean that the slogan "happy animals make good science" is a widely accepted statement? In general, scientists state that when housing conditions do not meet the demands of a particular species, one cannot expect reliable and reproducible results (Baumans, 2005), because animals that are under chronic stress or engaged in abnormal behaviours could be unsuitable research subjects. Animals from enriched housing conditions, on the other hand, are expected to be physiologically and psychologically more stable, they may be considered as more refined animal models, ensuring better scientific results (Bayne, 1996; Van de Weerd, 1996; Van de Weerd *et al*., 2002). In addition, the indications of increased welfare and decreased fearfulness in mice housed in more complex cages are expected to result in mice experiencing less stress when exposed to novel experiences during experimentation and reduced noise in experimental results (Olssen and Dahlborn. 2002).

Sherwin (2004) shows that standard, barren laboratory cages can influence the biology of laboratory animals, potentially violating one or more of the three conditions validity, reliability and reproducibility. He states, for example that laboratory animals from standard, barren cages often express heightened fearfulness, emotionality or reactivity. This could result in animals responding more vigorously to changes in the environment, for instance on behavioural tests. This increased reactivity, together with individual differences in responsiveness, most likely increase variability in data, thereby reducing the reproducibility of some studies (Sherwin, 2004). In addition, Garner and Mason (2002) state that abnormal behaviours might represent potential confound in scientific research. Due to the fact that standard, barren laboratory cages are related to a higher frequency of abnormal behaviours (Würbel *et al*., 1998; Würbel, 2001; Callard *et al*., 2000), many studies might be confounding in this way.

Furthermore, it seems that standard, barren cages do not always produce the quality of data we might expect. Hockley *et al*. (2002) concluded that environmental enriched mice mimic human disease more accurately, and that housing mice in a non-enriched environment cause a worsening in disease phenotype or neurological disorders. Thus, mice housed in standard, barren cages are unlikely a good model for human diseases (Hockley *et al*., 2002).

Concluding remark

Overall, it seems that Poole (1997) was right about his statement that 'unhappy' animals whose welfare is compromised are unsuitable subjects. Many scientists accept the idea that animal welfare can directly affects the scientific validity and repeatability of the data, due to the fact that compromised welfare can introduce unwanted variables into an experiment (e.g. abnormal behaviours), resulting in increased variance and non-reproducible data. In addition, the statement of Poole (1997) that 'happy' animals, whose welfare is not being compromised, are suitable subjects seems accepted. Improving animal welfare (i.e. through a good physical and social environment, good husbandry and good handling), reduces or prevents the occurrence of abnormal behaviour and physiology (Olssen and Dahlborn, 2002), and thus eliminates the confounding factor 'unhappiness' (i.e. the uncontrolled variables). Therefore it seems that using 'happy' animals whose welfare is not compromised results in good science due to a higher validity, reliability and reproducibility of the scientific data.

As discussed before, environmental enrichment can potentially improve animal welfare and thus the happiness of laboratory animals. Does this necessarily mean that the application of environmental enrichment results in good science, as proposed by Garner (2005)? This question brings us back to the two potential negative effects of environmental enrichment on scientific outcomes (i.e. the potential greater variation among data and the potential effect of reproducibility), which we also introduced in the last chapter. Does the application of environmental enrichment result in good science due to the improved welfare, or does the application result in poor science due to its potential greater variance among the data and its potential negative effect on reproductively? Both these question will be discussed in the following chapters.

Environmental enrichment and variation

The validity of animal experiments has been questioned on several grounds, one being that housing laboratory animals in barren cages results in impaired brain development, abnormal behaviour and physiology. This could potentially lead to artefacts in the results of animal experiments (Würbel, 2001; Garner, 2005). As discussed previously, this problem can be solved by making the cage environment more stimulating with the addition of environmental enrichment. However, despite the benefits of environmental enrichment for the welfare of laboratory animals, concerns have been raised that enriched housing might disrupt environmental standardization and reduce the precision and reproducibility of experimental results and hence validity (Gärtner, 1999; Wolfer *et al*., 2004). This could result in an increase in the number of laboratory animals needed in order to reach statistical significance (Eskola *et al*., 1999; Gärtner, 1999, Tsai *et al*., 2002).

This concern is based upon the assumption that a more complex enriched environment produces a higher diversity of behaviour within the home-cage and, therefore, increases the inter-individual variability in the responses of the laboratory animals to experimental treatments (Würbel and Garner, 2007). Several studies have empirically investigated this claim, resulting in conflicting findings; some supporting the claim (Eskola *et al*., 1999; Gärtner, 1999; Tsai *et al*., 2003) and others reporting no adverse affect on variation (Weerd *et al*., 2002; Augustsson *et al*., 2003; Wolfer *et al*., 2004; Lewejohann *et al*., 2006).

No adverse effect of environmental enrichment on variability

Van de Weerd *et al*. (2002) investigated the effects of enrichment on the results of a number of behavioural and physiological parameters in two routine testing procedures. They concluded that the enrichment did not influence the variability in any of the parameters measured. Van de Weerd *et al*. (2002) argued that this is probably due to the fact that animals housed in an enriched environment, can perform more of their species-specific behaviour, enabling them to cope better with novel and unexpected changes, thus showing an uniform response (Baumans, 1997).

Similar findings were found in several other studies. Augustsson *et al*. (2003) investigated if cage enrichment induced an effect on experimental mean values and on inter-individual variation. Different from all other studies, they choose to use three different statistical methods for analyzing variation with the assumption that these three methods might influence the interpretation of within-group variability. However, none of the methods showed any significant differences between standard and enriched conditions on variability in any of the parameters measured.

Lewejohann *et al*. (2006) applied systematic variation of housing conditions, laboratories and experimenters in their study, in order to test the influence of these variables on the experimental results of behavioural tests. They concluded that, in contrast to concerns that environmental enrichment might increase variability, the enrichment applied in their study (i.e. a mild form of enrichment consisting of a plastic inset and a wooden climbing frame) decreased variability (Lewejohann *et al*., 2006).

In addition, in 2004, Wolfer and colleagues showed that environmental enrichment did not increase individual variability in behavioural tests or the risk of obtaining conflicting data in replicate studies. They came to this conclusion by raising female mice of two inbred strains and their F1-hybrids in three different laboratories. The mice were housed in two different types of housing: (1) small, standard cages and (2) large, enriched cages. By the time their mice reached adulthood, they were subjected to four behavioural tests: elevated O-maze, open-field test, novel-object test and place navigation in the water maze. Each laboratory ran three independent replicates for each test. Wolfer and colleagues split the data by housing condition and calculated the proportion of variance for each replicate contributed by within-group variability and by laboratory X strain interactions. Then, they compared these proportions of variance between enriched and standard housing conditions (see figure 4). As shown in figure 4, the within-group variability contributed between 40% and 84% (i.e. an average of 60%)

to the total variance, while the contribution of laboratory X strain interactions (i.e. an average of 7.6%) was smaller and also less variable. Except for the behavioural test O-maze, the within-group variability was unaffected by enriched housing, and the enriched housing had no effect on the proportions of variance contributed by laboratory X strain interactions (Wolfer *et al*., 2004). Furthermore, the results indicated that enrichment did not increase the risk of obtaining conflicting results between the three different laboratories (Wolfer *et al*., 2004).

Thus, based on their results, Wolfer and colleagues (2004) concluded that environmental enrichment did not increase individual variability or the risk of obtaining conflicting data in replicate studies. Their findings indicate that the housing conditions of laboratory mice can be markedly improved without affecting the standardization of results. An important note for this study is that it was conducted with female mice, thus it remains to be seen whether their conclusions also apply to male mice. However, for female mice, environmental enrichment should improve their welfare without reducing the precision and reproducibility of the data derived from them (Wolfer *et al*., 2004).

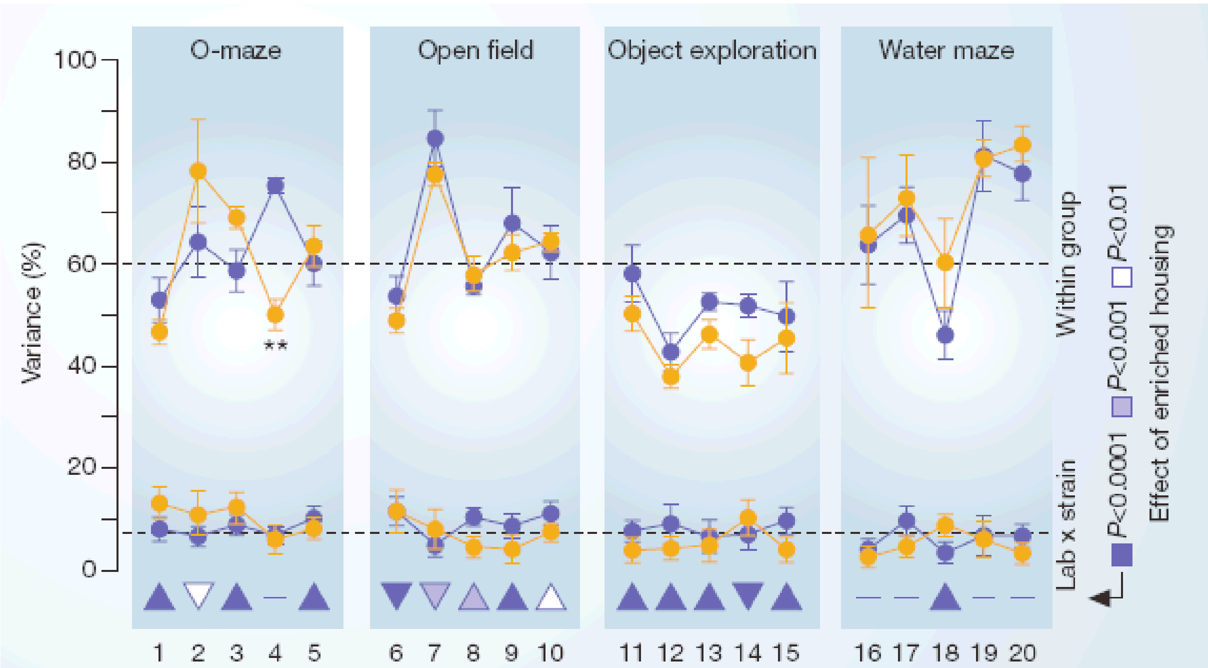


Figure 4. From Wolfer et al., 2004. Mean (± sem) proportion of variance in representative measures of four behavioural tests, contributed by within-group variability and laboratory X strain interactions, comparing female mice housed under standard (orange) and enriched (blue) conditions. Triangles indicate direction and significance of enrichment effects on each variable. Double asterisks, significant difference in individual variability.

Environmental enrichment increases variability

Tsai *et al*. (2002) performed an experiment for which three inbred mice strains were used. When the mice were 3 weeks old, they were separated randomly to enriched (i.e. containing a nest box, a wood bar for climbing and nesting material) or non-enriched cages. Body weights were recorded every week. Blood samples were collected at 14 weeks of age in order to analyze the white blood cells, red blood cells, haemoglobin and haematocrit. At 15 weeks of age, the mice were euthanized by CO2 in their home cages, and final body weight and organ weights were measured immediately.

Tsai and colleagues (2002) found that nearly all the test variables were not affected by environmental enrichment in their mean values. However, the enriched group showed different influences in variation. Depending on the strain of mice and the variables studied, environmental enrichment increased, decreased or had no effect on the variation. However, in general, they concluded that mice housed in enriched cages showed a higher variation for most measured parameters (Tsai *et al*., 2002).

In addition, Mering and colleagues (2001) evaluated the effects of the housing environment, including enrichment, cage type and group size, on the within-group variation and the number of laboratory animals needed to detect the chosen effect size. They concluded that it would be difficult to predict the effects of the housing environment on within-group variation, due to the fact that the effects seem to be time-, place-, animal-, and parameter-dependent (Mering *et al*., 2001). However, they pointed out that the effect of the housing environment on the experimental results should not be overlooked; "Although environmental enrichment is recommendable and it may even reduce the number of animals needed, it brings a new variable into research, which may complicate the evaluation of scientific data" (Mering *et al*., 2001).

Concluding remark

At the beginning of the discussion about whether environmental enrichment would lead to more variation in the results and hence increase the number of animals used, there were two groups having completely opposite thoughts. The first group, which strongly beliefs in the benefits of environmental enrichment, argued that animals housed in an enriched environment tend to be less reactive to stressful experimental situation, therefore, react with a more homogeneous response resulting in less variation (Baumans, 1997; van de Weerd *et al*., 2002). At the same time, the other group argued that a more complex enriched environment would produce a higher diversity of behaviour within the home-cage and, therefore, would increase the inter-individual variability in the responses of the laboratory animals to experimental treatments (Würbel and Garner, 2007). In order to find evidence for one of these claims, scientists started to investigate the influence of environmental enrichment on the variation of the results more than a decade ago.

Now, more then ten years later, the discussion has not yet ended, due to the fact that environmental enrichment has been reported to increase, decrease, or not affect variability, depending on the parameter studied (Eskola et al. 1999; Tsai et al. 2003; Van de Weerd *et al*., 2002). There are a few explanations for these inconclusive findings. First, the effect of environmental enrichment on the variation of the results could be parameter dependent, meaning that this influence varies between different species, different strains, different sex, different age, but also between the different types of environmental enrichment strategies. The second explanation is given by Augustsson and colleagues (2003). They state that in many laboratories, enrichment items are changed at irregular intervals and that, for example, the amount of nesting material differs between cages. This may result in a higher variation than if the use of environmental enrichment is standardized. The third, and last, possible explanation for these inconclusive findings, is that environmental enrichment could affect the reproducibility of experimental results, due to the fact that environmental enrichment potentially conflicts with standardization. This means that the use of environmental enrichment could potentially prevent that a certain result found in an experimental research, will be found again in a replicate study. Whether this is a valid explanation will be discussed in the next chapter.

Due to the inconclusive results and the three possible explanations, it is difficult to draw a clear conclusion on this subject. Therefore, instead of drawing conclusions, I would like to present my point of view in this discussion by assuming the worst case scenario. What if environmental enrichment indeed increases variability in the results and hence more animals are needed in order to gain the same statistical power, while at the same time, environmental enrichment enhances animal welfare? If this is the case, than the decision of providing environmental enrichment becomes difficult, because different aspects of the principles of refinement, reduction, and replacement ("the 3 Rs") are involved (Russell & Burch, 1959). The application of environmental enrichment to the animal's environment introduces refinement (i.e. refinement means that the degree of suffering is minimised; Russel, 2005). However, if the environmental enrichment also causes more variability in the study, and thus more animals are needed to achieve appropriate statistical power, then the goal of using fewer animals (i.e. reduction) is not achieved. Therefore, scientists need to balance the issues of enhanced animal welfare with the potential for reduced animal numbers used in the research (Baumans, 2005).

Thus, even when enrichment increases variation within the experimental study, it is important not to overstate the variation but instead, to balance the variation against the improved wellbeing of the animals. Although the goal of good science is to minimize all sources of variation in order to achieve valid and reliable results in animal-based research, the possible variation introduced by environmental enrichment might not be a negative factor because it allows the laboratory animals to express more of their species-specific behaviour, and that the experiment performed on a non-stressed and healthy animal (i.e. 'happy' animal; Poole, 1997) has indeed led to more reliable results (Young 2003).

In addition, when environmental enrichment increases variability, it could be argued that environmental enrichment should not be included. However, if environmental enrichment improves welfare of the animals to a reasonable extent it will most likely compensate for the increased number of animals used. From a utilitarian view, a better life for many may be considered preferable to a poorer life for the few (Sorensen *et al*., 2004; Hansen *et al*., 1998).

If scientists decide to provide environmental enrichment, there are two solutions with regard to its possible influence on the variability of the data. First, it would be advised to implement relatively simple environmental enrichment, because this will influence the variability less than a super-enriched complex cage (Baumans, 2005). Second, it is very important to describe the type of environmental enrichment sufficiently in the Material and Methods section of scientific publications to ensure the reproducibility of experimental results (Baumans, 2005).

Thus, in my view (and possibly from other scientists as well), implementing a simple environmental enrichment strategy in combination with the description of the type of environmental enrichment in the Material and Methods section, environmental enrichment should be applied to the cages of laboratory animals. It potentially improve animal welfare and thus result in 'happy' animals. And, as became clear from the previous chapter, using 'happy' animals will result in reliable conclusions because they do not have physiological and behavioural disabilities. Therefore, while scientist perform research in order to provide more evidence to end this discussion, environmental enrichment strategies should be provided, because rather a better life for many than a poorer life for the few (Sorensen *et al*., 2004; Hansen *et al*., 1998). And in addition, when enrichment does not influence variability or even decreases it, which is still possible, there is no reason for not introducing environmental enrichment and, thus, contributing to the welfare of laboratory animals (Van de Weerd *et al*., 2002).

Standardization and systematic environmental variation

Another way in which the validity of animal experiments has been challenged is by questioning the external validity of their results (Würbel and Garner, 2007). It is widely believed that environmental standardization is the best way to guarantee reproducible results in animal experiments. However, mounting evidence indicates that even subtle differences in laboratory (e.g. environmental enrichment) or test conditions can lead to conflicting test outcomes.

In 1999, Crabbe and colleagues performed a study in order to investigate the intra-laboratory reproducibility of findings. They addressed this problem of six mouse behaviours simultaneously in three laboratories using exactly the same inbred strains and one null mutant strain (Crabbe *et al*., 1999). The scientists in this study went to great lengths to try and reproduce laboratory-testing conditions across the three laboratories. Despite their efforts to equate laboratory environments, they did find disparate results across the laboratories. Crabbe and colleagues (1999) concluded that large strain differences are robust and unlikely influenced by site-specific interactions, but for behaviours with smaller genetic effects, there can be influences of environmental conditions specific to individual laboratories and even seemed to be idiosyncratic to a specific laboratory.

Conflicting findings between replicate studies, such as the study of Crabbe *et al*. (1999) have led to a debate about standardization from which two opposing views have emerged; (1) more rigorous environmental standardization will resolve the problem of poor reproducibility (Wahlsten, 2001; Öbrink and Rehbinder, 2000; van der Staay and Steckler, 2001, 2002), and (2) standardization itself is the cause of this problem (Würbel, 2000, 2002; Würbel and Garner, 2007). This debate contributes to the question whether environmental enrichment conflict with standardization and thus potential leads to less reproducible data.

Rigorous environmental standardization solves problem

Standardization is advised to ensure the consistency among scientists and comparability of data across different laboratories (van der Staay *et al*., 2009). There are, for instance, three factors which may substantially affect the results of scientific research: (1) the genetic composition of the animals used (e.g. outbred *versus* inbred strains), (2) the test situation (e.g. apparatus and protocol), and (3) the laboratory environment in which the animals are housed (e.g. the animal's environment and handling) (Wahlsten, 2001). If one or more of these three factors differs between replicate studies, the change of conflicting findings increases due to their substantially effect on the results of scientific research. This is exactly the argument on which the first group based their theory; results may differ both within and between laboratories due to the fact that some factors were not standardized, and rigorous standardization will resolve this problem of conflicting findings between replicate studies (Lewejohann *et al*., 2006; van der Staay and Steckler, 2001; 2002; Wahlsten, 2003).

Whereas standardization of the genetic composition of laboratory animals and the test situation is possible and desirable, standardization of laboratory environment seems impossible. That is, according to Wahlsten (2001), the reason why Crabbe and colleagues (1999) found differences in their results even after going through great lengths of standardization: "different experimenters at the three laboratories probably presented idiosyncratic arrays of odour cues and handled the mice somewhat differently". This influence of handling is exactly what Poole (1997) already mentioned. Most often, standardization of the laboratory environment results in a standardized housing environment of the laboratory animals. In other words, laboratory animals are housed in small standard barren cages, often in social isolation, in order to standardize their environment. It seems logic, because due to the lack of stimuli, structural features and complexity, every individual will experience this environment similarly. To illustrate, when laboratory animals are socially housed, the dominant and subordinate will experience this differently resulting in more variation (e.g. subordinates have a higher level of stress hormones, experience more fear and have a suppressed immune system;

Hucklebridge *et al.* 1976; Benton *et al.* 1978; Beden and Brain, 1984, 1985). Thus, standardization of the laboratory environment, including the housing conditions, could potentially result in less variation between the animals. However, there is also a negative effect of this standardization of the housing conditions, namely the development of abnormal behaviours and physiology due to the limited opportunities for the animals to perform species-specific behaviours (e.g. social behaviour, exploring, seeking shelter, foraging, gnawing and nest building; Baumans, 2005; Olssen and Dalhborn, 2002; Balcombe, 2006; Jennings *et al*., 1998; Latham and Mason, 2004). And, as mentioned before, behavioural and physiological disabilities can be considered confounding factors due to their uncontrolled variables (Poole, 1997). Thus, is standardization of the laboratory environment a valid method?

However, to guarantee reproducibility of experimental results, laboratory animal science textbooks advise experimenters to standardize the conditions of their experiments. According to Beynen and colleagues (2003), environmental standardization reduces variation in the data (i.e. within-experiment) and thereby maximizes test sensitivity. Higher test sensitivity allows a reduction of the sample size and thus the number of laboratory animals used. Therefore, standardization is promoted for ethical reasons (Festing, 2004a and 2004b). Furthermore, environmental standardization reduces variation between different laboratories (i.e. between-experiment) and thereby increases the reproducibility of the results (Beynen *et al*., 2003). Van der Staay (2006) argues that results are preliminary as long as investigators, preferably other than those who originally performed the investigations, have not confirmed the results. However, a problem with this standardized approach is that one cannot be sure that the findings are also valid under different conditions, i.e. whether the findings can be generalized (van der Staay and Steckler, 2002). Nonetheless, rigorous standardization is recommended because any limitation of generalization of results due to standardization is considered to be small (Beynen *et al*., 2003).

According to Van der Staay and Steckler (2001), many scientists have wrongly interpreted the results from the study of Crabbe *et al*. (1999), to demonstrate that standardization of test conditions does not increase the reliability of results. This conclusion, based on wrong interpretations, is premature (van der Staay and Steckler, 2001). They argue that even more rigorous standardization of all factors, that can potentially affect the experimental results, is needed to overcome these conflicting findings between replicate studies.

A very convincing example of the necessity to strictly standardise test conditions has been given by Crestani and colleagues (2000). Two groups (Crestani *et al*., 2000; McKernan *et al*., 2000; Rudolph *et al*., 1999) were able to show the same effects of a mutation of the a1GABAA receptor subtype under comparable testing conditions. However, initially the two lines of mice used by both groups seemed to differ in their drug-induced behaviour, although both had been constructed with the same point mutation, but Crestani *et al*. (2000) reported that the discrepancies were caused by differences in the behavioural protocols used by the two groups and not by differences in the mouse lines. Thus, under comparable test conditions, there are no apparent behavioural differences between the mice generated by the two laboratories (Rudolph *et al*., 1999; McKernan *et al*., 2000). This shows that taking environmental and technical details of test procedures into account may help to resolve inter-laboratory differences in results (Crestani *et al*., 2000). This must be considered as an example in favour of standardised testing conditions (van der Staay and Steckler, 2002).

Standardization is the cause of the problem

As mentioned previously, in laboratory animal science it is advised to standardize environmental conditions in order to reduce variation in the data and increase reproducibility of results (Beynen *et al*., 2003). Indeed, standardization within a laboratory may reduce variation in the data and increase reproducibility of the results within that laboratory (Würbel and Garner, 2007). However, many environmental factors resist between-laboratory standardization (e.g. staff, room architecture and noise levels; Crabbe *et al*., 1999), resulting in laboratories standardizing to different local condition

(Wahlsten, 2001; Würbel, 2002). If a treatment response varies with local conditions, the standardization to different local conditions in laboratories will lead to results from different laboratories becoming more and more distinct (Würbel and Garner, 2007), and thus decreases the reproducibility of the results between laboratories.

The reproducibility of results is used as a measure of external validity. External validity stands for the ability to apply the results to other conditions (i.e. environmental contexts), populations or species (Campbell, 1957) and refers to the robustness of the results outside the circumstances in which it was established (Guala, 2003) and thus defines the extent to which a result can be generalized (Richter *et al*., 2009). Thus, if a result can be reproduced in a second experiment, either in the same or a different laboratory, the result is confirmed to be robust and externally valid. Those who recommend standardization to increase reproducibility of results (Beynen *et al*., 2003) state that experimental results only hold for the conditions under which the experiment has been carried out, and that therefore, the results obtained under standardized conditions can not be generalized (Beynen *et al*., 2003).

Thus it is acknowledged that increased reproducibility through standardization is at the expense of external validity (Würbel, 2000). This is referred to as standardization fallacy and means that a result is highly reproducible under highly standardized conditions, but can be poorly generalized to other conditions (Würbel, 2000). Nonetheless, rigorous standardization is recommended because any limitation of generalization of results due to standardization is considered to be very small and thus negligible (Beynen *et al*., 2003; Wahlsten, 2001).

Several studies have indeed shown that despite extraordinary efforts to standardize environmental test conditions between laboratories, the experimental results were poorly reproducible. Wahlsten *et al*. (2003) decided to address the question of reproducibility of results in different laboratories systematically, by testing the resilience of genotypic influences, using mice of eight genotypes simultaneously in three laboratories on a battery of six simple behaviours (Crabbe *et al*., 1999). They standardized apparatus, test protocols, and other environmental variables to minimize non-genetic sources of variability. The results of their study indicate that standardization of the test situation does not guarantee identical results in different laboratories, because of large effects of laboratory environments.

In another study, performed by Lewejohann *et al*. (2006), systematic variation of housing conditions, laboratories and experimenters was applied in order to test the influence of these variables on the outcome of behavioural tests. In spite of their extensive protocol standardizing laboratory environment (i.e. light intensity, light/dark cycle, temperature and test apparatuses), animal maintenance and testing procedures between laboratories and experimenters, significant differences between different laboratories and different experimenters were found.

Wahlsten and colleagues (2006) wondered whether conducting the same experiment in two laboratories or repeating a classical study many years later, would obtain the same results. They tried to answer this question by selecting historical comparisons studies based on previously published reports of tests given to inbred strains from The Jackson Laboratory. They concluded that the systematic comparison of recent and classical data sets was not sufficient to draw conclusions about what kinds of behaviour would be most and least robust across laboratories. However, they suggested that behaviours closely associated with sensory input and motor output would be less affected by minor variants in the laboratory environment, whereas behaviours related to emotional and social processes would be more affected (Wahlsten *et al*., 2006).

Thus, both theoretical considerations and empirical evidence indicate that the concept of environmental standardization seriously limits external validity of animal experiments and therefore decreases reproducibility of the results (Würbel and Garner, 2007). Thus those who argue that environmental standardization reduces animal numbers by reducing variation in the data simply ignore that additional animals will be needed in replicate studies to test whether the results are actually robust against even minor variations in the environment (i.e. testing externally valid) (Würbel and Garner, 2007).

Systematic environmental variation

Empirical evidence indicates that subtle differences between laboratory, housing or test conditions can lead to conflicting test outcomes and raise the question as to whether standardization is the appropriate approach to resolve this problem (Richter *et al.*, 2009). In order to gain reproducible results across laboratories, the experimental results should have to be applicable to at least the range of environmental conditions covered by laboratory differences (Richter *et al.*, 2009). In addition, for the results to be relevant they should be reasonably robust against some variation in the environment (Würbel and Garner, 2007).

Richter and colleagues (2009) state that due to standardization, animals within laboratories will be more homogenous than animals between laboratories, due to the resistance to between-laboratory standardization of environmental factors. Therefore they propose systematic and controlled environmental variation, which should improve the reproducibility of results, because systematic environmental variation renders the animals within experiments more heterogeneous, increasing the external validity of the results, without confounding the variation due to treatment (Richter *et al.*, 2009).

The proposal of systematic environmental variation by Richter and colleagues (2009) was based on the results of their study. They used previously published data (Wolfer, 2004) from female mice of two inbred strains (C57BL/6J, DBA/2) and one hybrid strain (B6D2F1) that had been tested for behavioural strain differences in a multi-laboratory study involving three different laboratories. Each laboratory had three independent batches of mice that were housed in either enriched or unenriched cages. These female mice of three strains, distributed across three laboratories, three batches per laboratory and two housing conditions, were allocated to 18 standardized and 18 heterogenized (i.e. made possible through systematic environmental variation) replicate cohorts. Then they compared the variance and the rate of 'false positive' and 'false negative' results in the standardized and heterogeneous group to the pooled findings.

They found that the variance between replicate experiments was greater in standardized replicates compared to heterogenized replicates, indicating that standardization results in lower reproducibility of the results. Furthermore, they found that standardized replicates produced 9.4% 'false positive' results, compared to 1.3% among heterogenized replicates, indicating that standardization increases the rate of idiosyncratic results. In addition, the rate of 'false positive' results in standardized replicates, in contrast to heterogenized replicates, was significantly higher than expected by chance alone, indicating that standardization may introduce a systematic source of 'false positive' results.

Based on these results, Richter and colleagues (2009) concluded that environmental standardization introduced a systematic source of idiosyncratic results above chance alone and thus cannot guarantee reproducible results. The increased risk for spurious results, results in the fact that standardization may create scientific uncertainty in many areas of laboratory research. This will lead to the need for replicate studies, which causes economic costs and, with respect to animal research, undermines the ethical goal of reducing animal use by increasing test sensitivity. In addition, their findings demonstrate that the reproducibility of results may be improved and spurious results avoided, by introducing adequate environmental heterogenization into the experimental design.

However, Paylor (2009) challenged their findings. He claimed that the findings of Richter and colleagues (2009) were based upon retrospective analysis and that heterogenization was logistically unfeasible. Paylor (2009) states that there may be some disagreement about how the particular behavioural responses were selected and about their choice to use only the extremes of minimized (standardized) and maximized (heterogeneous) environment variation groups. However, their approach can be considered reasonable and the implications should not be taken lightly. According to Paylor (2009), the most logical next step would be to empirically test the hypothesis. This was (although not with those intents) already done to some extent in the study of Crabbe *et al*. (1999). Unfortunately, this study (Crabbe *et al*., 1999) did not attempt to lighten the standardization in order to determine whether a heterogeneous approach would have yielded more reproducible results. This, however, was done by Richter and colleagues (2010), who tested and confirmed their hypothesis empirically, showing that even simple heterogenization within experiments would improve reproducibility between experiments (Richter *et al.*, 2010).

In this experiment, they allocated 256 female mice of two inbred strains (C57BL/6 and BALB/c) to four standardized and four heterogenized replicate experiments, and examined strain differences in three behavioural tests (free exploration, open field and novel object). To mimic different laboratories, they varied experimental conditions between replicate experiments. They found that the strain differences were relatively consistent among heterogenized experiments, while they varied considerably between standardized experiments. In addition, the between-experiment variation was lower in the heterogenized design, indicating better reproducibility (Richter *et al.*, 2010). Furthermore, they found that the effect of the 'strain' was stable in the four heterogenized experiments, while the outcomes of the four standardized experiments were highly variable, suggesting that within-experiment standardization increased test sensitivity at the expense of reproducibility.

Based on their results, Richter and colleagues (2010) concluded that their findings empirically confirmed that standardized experiments could generate spurious results by increasing test sensitivity at the expense of external validity. They also concluded that even simple forms of heterogenization could render study populations sufficiently heterogeneous to guarantee robust results across the unavoidable variation between experiments (Richter *et al.*, 2010).

How to apply systematic environmental standardization?

Unfortunately, both studies of Richter *et al.* (2009; 2010) failed to provide direct guidelines for the scientists who would be interested in trying to use a more heterogeneous environment for their behavioural studies. In the essence, they argue that to generate results that are most likely going to be reproducible in other laboratories, the strategies to standardize environmental conditions in an experiment should be minimized (Paylor, 2009). According to Richter and colleagues (2009), methods are needed for within-laboratory heterogenization resulting in populations that better represent the range of environmental variation between laboratories. The age of the animals and various aspects of their housing conditions (for example, cage size, type of enrichment and group size) are promising variables in this respect (Richter *et al.*, 2009). An example of this within-laboratory heterogenization is shown in figure 5. Next to the differences of the animals themselves (e.g. age), the housing condition of the cages in which the animals are housed are heterogenized as well.

An important, critical note put forth by Richter *et al.* (2009), is that systematic environmental variation would improve the reproducibility of results, as long as the animals are 'matched'. This means that for each treatment animal, a control animal is selected from the same microenvironment.

According to Würbel and Garner (2007) simple solutions do exist in order to achieve systematic environmental variation; using randomized block designs (figure 6). Randomized block designs allow the introduction of environmental variation in a systematic and controlled way, without the need for larger sample sizes, while at the same time increasing precision and statistical power (Beynen *et al.*, 2003). The increased precision and statistical power with randomized block designs are due to the fact that inter-individual variation within each block is normally smaller than overall variation in an unblocked design and due to the fact that comparing treatments within blocks only eliminates the between-block variation (Beynen *et al.*, 2003).

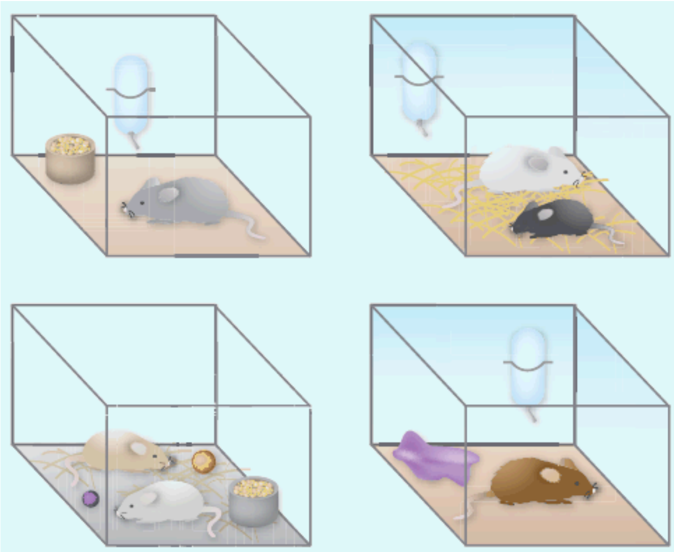


Figure 5. From Paylor (2009). Heterogeneous conditions, minimized standardization, lead to more reproducible behavioural results. Both the animals themselves (e.g. age) and the housing conditions of the animals are heterogenized.

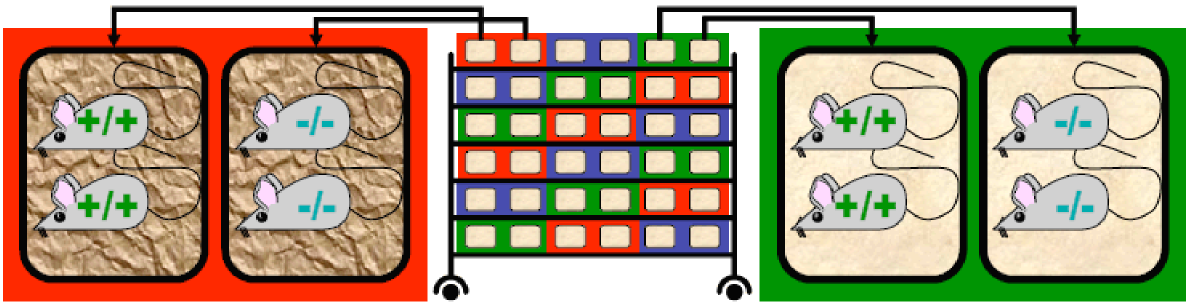


Figure 6. From Würbel and Garner (2007). A randomized block design for an experiment in which genotype is considered the treatment. The animals are housed in three different housing systems (red, blue, green) to introduce systematic environmental variation. The three housing conditions differ from each other. As shown, housing variants red and green e.g. differ in whether they do (red) or do not (green) contain shelter in addition to nesting material. Variant blue (not shown here) might e.g. be larger cages or involve pre-test handling. The combination of cage and housing variant represents the blocking factor in the statistical design. In principle, however, each coloured cell of two paired cages might represent a slightly different environment. Using cell as blocking factor controls for the environmental variation between the cells and increases external validity of the results without inflating sample size (Würbel and Garner, 2007).

Concluding remark

The decision to rigorously standardize or to loosen experimental control with systematic environmental variation depends on the questions a study tries to answer and should be based on the specific needs to meet the intended goal (van der Staay *et al.*, accepted article). Most often, the goal of experiments and using animal models is to generalize the findings to humans and/or a species other than the one studied, or in the same species to conditions different from those under which the study was performed (Arndt & Surjo 2001, van der Staay 2006). As discussed above, results obtained in highly standardized studies are only valid under those exact conditions. Thus, if the goal of a study is to generalize to a larger population, one might heterogenize environmental conditions with systematic environmental variation. However, if the effects are subtle, the chance of false negative findings will increase and the effects may remain undetected when tested under heterogeneous environmental conditions. This might be due to some environmental conditions promoting, while others suppressing the expression of a specific behavioural phenotype, or the effects become unnoticed in the noise caused by uncontrolled factors (Calisi & Bentley 2009).

On the other hand, a strictly standardized study might fail to detect the effects, because the environmental conditions may not promote the expression of a particular phenotype. However, in situations in which false negatives can be counterproductive or even dangerous (e.g. in toxicological studies and in drug safety testing), experiments based on standardization may provide more useful information (van der Staay *et al*., accepted article).

Thus in conclusion, systematic environmental variation should not be applied in proof of principle studies, considering the higher risk of false negative findings, unless the expected effects are large and robust (van der Staay *et al*., accepted article).

How can this all together be translated to the discussion whether happy animals make good science? To begin with, standardization does not seem to fit in with the happiness of laboratory animals. As discussed in the beginning of this chapter, standardization of the laboratory environment results in a standardized housing environment of the laboratory animals. In other words, laboratory animals are housed in small standard barren cages, often in social isolation, in order to standardize their environment. These housing conditions can result in the development of abnormal behaviours and physiology due to the limited opportunities for the animals to perform species-specific behaviours (e.g. social behaviour, exploring, seeking shelter, foraging, gnawing and nest building; Baumans, 2005; Olssen and Dalhborn, 2002; Balcombe, 2006; Jennings *et al*., 1998; Latham and Mason, 2004). And, these behavioural and physiological disabilities can be considered confounding factors due to their uncontrolled variables (Poole, 1997).

Furthermore, supporters of rigorous standardization already point out that the laboratory environment cannot be completely standardized (in comparison to the genetic composition and test situation). In my opinion it would make sense to allow this unavoidable variation in a controlled and systematic way, which is the case with systematic environmental variation. In addition, systematic environmental variation is a method which applies within-laboratory heterogenization, resulting in populations that represent the range of environmental variation between laboratories, and hence improves the reproducibility of experimental data. Thus, the external validity of the results is increased, because the results can be applied in different environmental contexts. Furthermore, systematic environmental variation allows the provision of environmental enrichment, resulting in 'happy' animals and hence in reliable conclusions (Poole, 1997).

Thus, systematic environmental variation should be applied in order to guarantee good science and the happiness of the laboratory animals. However, it would still be advisable to use standardization rather than systematic environmental variation in proof of principle studies, considering the higher risk of false negative findings of with systematic environmental variation, unless the expected effects are large and robust (van der Staay *et al*., accepted article).

Discussion

Important findings

With his statement "happy animals make good science", Poole (1997) argues that scientific research should be based upon healthy animals that have a normal physiology and express normal behaviour, and that the use of animals whose welfare is compromised (i.e. animals that have physiological and behavioural disabilities), results in unreliable conclusion due to the fact that these disabilities can be considered a confounding factor. Thus "happy animals make good science' refers to the fact that animals whose welfare is not being compromised are suitable subjects for science, while animals whose welfare is being compromised are unsuitable subjects (Poole, 1997). This thesis discussed several questions in order to evaluate the influence of environmental enrichment on the statement of Poole (1997), "happy animals make good science".

In order to evaluating the influence of environmental enrichment, it was important that each step was evaluated separately, before drawing conclusion. The important findings of this evaluation are summarized. Animal welfare is the physical and physiological wellbeing of animals. It is believed that animal welfare is present when an animal can reliably predict and control relevant events by means of species-specific behaviours (Wiepkema and Koolhaas, 1993). If relevant events become permanently unpredictable and uncontrollable, it will results in an environment which is beyond the control of animals; they cannot adapt to and cope adequately with their environment and their welfare is compromised. In this situation, chronic stress symptoms become visible: abnormal behaviour develop (e.g. stereotypies), and the baseline level of physiological measures becomes chronically elevated (e.g. plasma corticosteroid levels) (Wiepkema and Koolhaas, 1993). The housing systems for laboratory animals have been designed according to hygienic, economic and ergonomic requirements to provide standardized conditions with little or no consideration for animal welfare (Baumans, 2004, 2005; Olsson *et al.*, 2003). This type of housing condition (i.e. small, standard, barren cages) provides limited stimuli, prevent laboratory animals from exerting control over their situation, and gives limited opportunities for the animals to perform species-specific behaviours (e.g. social behaviour, exploring, seeking shelter, foraging, gnawing and nest building) (Baumans, 2005; Olssen and Dalhborn, 2002; Balcombe, 2006; Jennings *et al.*, 1998; Latham and Mason, 2004).

Thus, standard laboratory cages compromise the welfare of laboratory animals and results in animals with behavioural and physiological disabilities. The use of these animals for scientific research would not be appropriate according to Poole, due to the fact that these animals can be considered confounding factors due to the many uncontrollable variables in their behavioural and physiological disabilities. He based this upon several studies which investigated the influence of social factors, the physical environment and the handling on the happiness of laboratory animals. If animals are not housed in the right social condition, if animals have to deal with uncontrollable, unpredictable husbandry events, and if animals are not treated kindly and gently, the welfare of these animals can be compromised resulting in animals with behavioural and physiological disabilities. Thus according to Poole (1997) scientists are obligated to do everything practicable to ensure the happiness of laboratory animals, due to the fact that 'happy' animals are healthy and have a normal physiology and behaviour. These animals are able to cope with the stressors to which they are subjected, and will not be in pain, or distressed and resist disease. If these happy animals receive a good and simple treatment, stress will be even more reduced and animals become confident, cooperative and easily handled. Therefore, 'happy' animals are suitable subjects. Many scientists accept this idea that animal welfare can directly affects the scientific validity and repeatability of the data, due to the fact that compromised welfare can introduce unwanted variables into an experiment (e.g. abnormal behaviours), resulting in increased variance and non-reproducible data. In addition, improving animal welfare, reduces or prevents the occurrence of abnormal behaviour and physiology (Olssen and Dahlborn, 2002), and thus eliminates the confounding factor 'unhappiness' (i.e. the uncontrolled variables).

One way to reduce or prevent the occurrence of abnormal behaviour and physiology, and potentially increase animal welfare, is the application of environmental enrichment. The provision of environmental enrichment, results in an increased complexity of an enclosure in order to provide the opportunity for animals to perform species-specific behaviours, with the goal of improving animal welfare (Young, 2003; Baumans, 2005). However, although intuition and even direct observation indicate that environmental enrichment promotes species-specific behaviours, data showing that rodents benefit from these changes are often inconclusive (Benefiel *et al*., 2005; Olssen and Dahlborn, 2002). This means that the results (i.e. benefit for animal welfare) often vary between different species, different strains, different sex, different age, etc.

However, despite the lack of scientifically evidence and scientifically based guidelines, environmental enrichment is a widely accepted practice. In addition, many recent studies on rodent welfare support the view that environmental enrichment is by definition a good thing, without evidence that it is essential for animal welfare or consideration of its possible effects on experimental outcomes (Benefiel *et al*., 2005). Considering the effects environment enrichment can have on the physiology, brain structure and function, and the behaviour of animals, it makes sense that scientists have concerns about how these changes may affect experimental data and the validity of research results (Benefiel *et al*., 2005).

Studies have reported that environmental enrichment has the potential to increase, decrease, or not affect variability, depending on the parameter studied (Eskola et al. 1999; Tsai et al. 2003; Van de Weerd *et al*., 2002). Thus, the data on this matter is inconclusive. However, it is important to realise that when enrichment increases variation, the variation should be balanced against the improved welfare of the animals. Some scientist argue that when environmental enrichment increases variability, it should not be included. However, if environmental enrichment improves welfare of the animals to a reasonable extent it will most likely compensate for the increased number of animals used.

There was (and still is) another concern about the application of environmental enrichment, namely the fact that environmental enrichment compromises standardization. Standardization is advised to ensure the consistency among scientists and comparability of data across different laboratories (van der Staay *et al*., 2009). There are, three factors which may substantially affect the results of scientific research and should be standardized in order to prevent any effect: (1) the genetic composition of the animals used, (2) the test situation, and (3) the laboratory environment in which the animals are housed (Wahlsten, 2001).

In order to standardize the housing environment of laboratory animals, the animals are housed in small standard barren cages, often in social isolation. Due to the lack of stimuli, structural features and complexity, every individual will experience this environment similarly (e.g. no differences in stress hormone levels, fear expression and immune response; Hucklebridge *et al.* 1976; Benton *et al.* 1978; Beden and Brain, 1984, 1985). Thus, standardization of the laboratory environment, including the housing conditions, could potentially result in less variance between the animals. Whether this is true and evidence based, will be discussed later. Thus, scientists who strongly believe in standardization, think that environmental enrichment can disrupt standardization and result in non reproducible data.

However, conflicting findings between replicate studies, such as the study of Crabbe *et al*. (1999) have led to a debate about standardization from which two opposing views have emerged; (1) more rigorous environmental standardization will resolve the problem of poor reproducibility (Wahlsten, 2001; Öbrink and Rehbinder, 2000; van der Staay and Steckler, 2001, 2002), and (2) standardization itself is the cause of this problem (Würbel, 2000, 2002; Würbel and Garner, 2007). Supporters of rigorous standardization already point out that the laboratory environment cannot be completely standardized (in comparison to the genetic composition and test situation). It does make sense to allow this unavoidable variation in a controlled as systematic way, which is the case with systematic environmental variation. In addition, systematic environmental variation is a method which applies within-laboratory heterogenization, resulting in populations that represent the range of environmental variation between laboratories, and hence improves the reproducibility of experimental data. Thus, the external validity of the results is increased, because the results can be applied in different environmental contexts.

Future work

In this evaluation about the influence of environmental enrichment on the statement "happy animals make good science", a few findings resulted in even more questions. As already discussed in the last chapter, therefore I will not go into detail, it is argued that standardization decreases variation. It sounds intuitively right that standardizing the genetic composition of the laboratory animals and the experimental procedures can result in less variation. However, is this statement evidence based? Especially, regarding the standardization of the laboratory environment, standardization could potentially also result in increased variation (i.e. standard laboratory cages result in abnormal behaviour and physiology, see p. 29). What makes this even more interesting, standardization is widely accepted and scientists refer to it as a good method to decrease variability and thus number of animals used. But, if this is true, how come that so many replicate studies result in conflicting findings? Thus, it would be good to conduct research on this topic.

Another discussible subject is environmental enrichment and its influence on animal welfare and scientific results. Both these influences of environmental enrichment remain inconclusive, meaning that the results (i.e. benefit for animal welfare or its influence on variance) often vary between different species, different strains, different sex, different age, and between different types of environmental enrichment strategies. However, despite the lack of scientifically evidence and scientifically based guidelines, environmental enrichment is a widely accepted practice. Many recent studies on rodent welfare support the view that environmental enrichment is by definition a good thing, without evidence that it is essential for animal welfare or consideration of its possible effects on experimental outcomes. I think that this in particular, deserves more attention and a clear answer. Thus, more research should be conducted on this topic as well.

Applying environmental enrichment is largely based on what was intuitively perceived as important, and there is a need for more research to determine whether this is indeed evidence based. There are a few possible steps to evaluate the effect of environmental enrichment on animal welfare and experimental results (Olsson and Dahlborn, 2002). First, it would be a good idea to evaluate the effect of environmental enrichments on welfare related parameters (e.g. behaviour in the home cage, reaction to handling and novelty). This evaluation should be conducted with a sufficient number of animals of both sexes, various ages and at least from two strains. This is needed in order to determine whether the effect of environmental enrichment is parameter dependent. It is also important that effects of environmental enrichment are studied over a longer period, so that the recorded effect is not only a novelty effect.

If this evaluation results in the conclusion that certain environmental enrichment strategies improve animal welfare, a second experiment is needed. In this experiment, the same environmental enrichment strategies are evaluated, but this time for its effect on parameters relevant to the experiments (e.g. baseline physiology, immunology parameters or reactions in behavioural tests). These two types of experiment will provide a better view of the affect of environmental enrichment on the animal welfare and on experimental results.

Conclusion

Now that all important findings of the different evaluation steps are summed up and certain things discussed, one should be able to draw conclusions about the influence of environmental enrichment on the statement "happy animals make good science". To start with, the statement "happy animals make good science" from Poole (1997) is evidence based, and even 13 years after, scientists agree with his idea. Thus scientist should do everything practicable to ensure the happiness (i.e. welfare) of their laboratory animals.

One way to ensure the happiness of animals is the application of environmental enrichment. It has been shown that animals have a great preference for certain environmental enrichment strategies and are even willing to work for it. In addition, some environmental enrichment strategies reduce or prevent the occurrence of abnormal behaviour and physiology. Despite the inconclusive results (i.e. the beneficial effects of environmental enrichment are parameter dependent) no direct negative effects have been found on the welfare of animals. Therefore, environmental enrichment is considered to be a strategy that tackles the underlying causes of abnormal behaviour and physiology, and results in improved welfare.

Whether environmental enrichment results in more variation in the experimental data remains unclear, because studies have reported that environmental enrichment has the potential to increase, decrease, or not affect variability, depending on the parameter studied. However, in my opinion, if environmental enrichment improves welfare of the animals to a reasonable extent it will most likely compensate for the increased number of animals used due to its potential increase in variance. I agree with the utilitarian view; a better life for many may be considered preferable to a poorer life for the few.

These thoughts result in the fact that I personally believe that systematic environmental variation should be applied in laboratory animal science, because it allows the provision of environmental enrichment, resulting in 'happy' animals and in reliable conclusions. Furthermore, reproducibility and external validity is also increased. However, systematic environmental variation should not be applied in proof of principle studies, considering the higher risk of false negative findings, unless the expected effects are large and robust.

Thus, in conclusion, if environmental enrichment is applied according to the systematic environmental variation method, it will result in happy animals and in good science.

References

Arndt, S.S. and Surjo, D. (2001). Methods for the behavioural phenotyping of mouse mutants. How to keep the overview. *Behavioural Brain Research*, 125, 39-42.

Augustsson, H., van de Weerd, H.A., Kruitwagen, C.L.J.J. and Baumans, V. (2003). Effect of enrichment on variation and results in the light/dark test. *Laboratory Animals*, 37, 328-340.

Balcombe, J.P., (2006). Laboratory environments and rodents' behavioural needs: a review. *Laboratory Animals*, 40, 217–235.

Balcombe, J. (2010). Laboratory Rodent Welfare: Thinking outside the cage. *Journal of applief animal welfare science*, 13, 77-88.

Baumans, V. (1997). Environmental enrichment: Practical applications. In: Van Zutphen LFM, Balls M, eds. Animal Alternatives, Welfare and Ethics. Amsterdam: Elsevier BV. p 187-191.

Baumans, V. (2000). Environmental Enrichment: A right of rodents! In: Balls M, Van Zeller A-M, Halder M, eds. Progress in the Reduction, Refinement and Replacement of Animal Experimentation. Amsterdam: Elsevier BV. p 1251-1255.

Baumans, V. (2004). The welfare of laboratory mice. In: Kaliste E, ed. The Welfare of Laboratory Animals. Dordrecht: Kluwer Academic Publishers. p 119-152.

Baumans, V. (2005). Environmental enrichment for laboratory rodents and rabbits: Rodent, Rabbit, and Research Requirements. *ILAR Journal*, 46, 162-170.

Beden, S.N. and Brain, P.F. (1984). Effects of attack-related stress on the primary immune responses to sheep red blood cells in castrated mice. *IRCS Medical Science*, 12, 675.

Beden, S.N. and Brain, P.F. (1985). The primary immune responses to sheep red blood cells in mice of differing social rank or from individual housing. *IRCS Medical Science*, 13, 364-365.

Benton, D., Goldsmith, J.F., Gamal-el Din, L., Brain, P.F. and Hucklebridge, F.H. (1978). Adrenal activity in isolated mice and mice of different social status. *Physiology and Behaviour*, 20, 459-464.

Beynen, A.C., Gärtner, K. and van Zutphen, L.F.M. (2003). Standardization of animal experimentation In *Principles of Laboratory Animal Science* (eds., van Zutphen, L.F.M., Baumans, V. & Beynen, A.C.) 103–110 (Elsevier, Amsterdam).

Blecha, F., Kelley, K.W., Satterlee, D.G. (1982). Adrenal involvement in the expression of delayed-type hypersensitivity to SRBC and contact sensitivity to DNFB in stressed mice. *Proceedings of the Societyfor Experimental Biology and Medicine*, 169, 257.

Brain, P.F. (1990). Stress in agonistic contexts in rodents.In: *Stress in Domestic Animals* (Dantzer R, Zayan R, eds). Dortrecht: Kluwer Academic, pp 74-85.
Benefiel, A.C., Dong, W.K. and Greenough, W.T. (2005). Mandatory "enriched" housing of laboratory animals: the need for evidence based evaluation. *ILAR journal*, 46, 2, 97-105.

Broom, D.M. (1996). Animal welfare defined in terms of attempts to cope with the environment. *Acta Agriculture Scandinavica Section A, Animal Science, Supplement*, 27, 22-28.

Callard, M.D., Bursten, S.N. and Price, E.O. (2000). Repetitive backflipping behaviour in captive roof rats (*Rattus rattus*) and the effects of cage enrichment. *Animal Welfare*, 9, 139-152.

Calisi, R.M. and Bentley, G.E. (2009). Lab and field experiments: are they the same animal? *Hormones and Behaviour*, 56, 1-10.

Campbell, D.T. (1957). Factors relevant to the validity of experiments in social settings. *Psychological Bullitin*, 54, 297–312.

Chapillon, P., Manneché, C., Belzung, C. and Caston, J. (1999). Rearing enrichment in two inbred strains of mice: Effects on emotional reactivity. *Behaviour Genetics*, 29, 41-46.

Cooper, J.J., Ödberg, F. and Nicol, C.J. (1996). Limitations on the effectiveness of environmental improvement in reducing stereotypic behaviour in bank voles (Clethrionomys glareolus). *Applied Animal Behaviour Science*, 48, 237-248.

Crabbe, J.C., Wahlsten, D. and Dudek, B.C. (1999). Genetics of mouse behaviour: Interactions with laboratory environment. *Science*, 284, 1670-1672.

Crestani, F., Martin, J.R., Möhler, H. and Rudolph, U. (2000). Resolving differences in GABAA receptor mutant mouse studies. *Nature Neuroscience*, 3, 1059.

Dahlborn, K., van Gils, B.A.A., van de Weerd, H.A., van Dijk, J.E. and Baumans, V. (1996). Evaluation of long-term environmental enrichment in the mouse. *Scandinavian Journal of Laboratory Animal Science*, 23, 97-106.

Damon, E.G., Eidson, A.F., Hobbs, C.H. and Hahn, F. (1986). Effect of acclimation on nephrotoxic response *of* rats to uranium. *Laboratory Animal Science*, 36, 24-27.

Dawkins, M.S. (1998). Evolution and animal welfare. *The Quarterly Review of Biology*, 73, 305-328.

De Greeve, P.C.M. and De Leeuw, W.A. (2003). Wettelijke aspecten van dierproeven. In: van Zutphen, L.F.M., Baumans, V. and Beynen, A.C., eds. Handboek proefdierkunde. Proefdieren, dierproeven, alternatieven en ethiek. 4th edition, Elsevier gezondheidszorg.

Ellis, S. (2009). Environmental enrichment. Practical strategies for improving feline welfare. *Journal of Feline Medicine and Surgery*, 11, 901-912

Eskola, S. and Kaliste-Korhonen, E. (1999). Aspen wood-wool is preferred as a resting place, but does not affect intracage fighting of male BALB/c and C57BL/6J mice. *Laboratory Animals*, 33, 108-121.

Eskola, S., Lauhikari, M., Voipio, H.M., Laitinen, M. and Nevalainen, T. (1999). Environmental enrichment may alter the number of rats needed to achieve statistical significance. *Scandinavian Journal of Animal Science*, 26, 134–44.

Festing, M.F.W. (2004a). Good experimental design and statistics can save animals, but how can it be promoted? *Alternative Laboratory Animals*, 32, 133–135.

Festing, M.F.W. (2004b). Refinement and reduction through the control of variation. *Alternative Laboratory Animals*, 32, 259–263.

Garner, J.P. (2005). Stereotypies and other abnormal behaviours: potential impact on validity, reliability, and replicability of scientific outcomes. *ILAR journal*, 46, 2, 106-117.

Garner, J.P. and Mason, G.J. (2002). Evidence for a relationship between cage stereotypies and behavioural disinhibition in laboratory rodents. *Behavioural Brain Research*, 136, 83-92.

Garner, J.P., Mason, G. and Smith, R. (2003). Stereotypic route-tracing in experimentally-caged songbirds correlates with general behavioural disinhibition. *Animal Behaviour,* 66, 711-727.

Garner, J.P., Weisker, S.M., Dufour, B. and Mench, J.A. (2004). Barbering (fur and whisker trimming) by laboratory mice as a model of human trichotillomania and obsessive-compulsive spectrum disorders. *Comp Med*, 54,216-224.

Gärtner, K. (1999). Cage enrichment occasional increases deviation of quantitative traits. *Proceedings o the International Joint Meeting 12th ICLAS General Assembly and Conference & 7th FELASA Symposium*, 207-210.

Guala, F. (2003). Experimental localism and external validity. *Philosophy of Science*, 70, 1195– 1205.

Hockley, E., Cordery, P.M., Woodman, B. Mahal, A., van Dellen, A., Blakemore, C., Lewis, C.M., Hannan, A.J. and Bates, G.P. (2002). Environmental enrichments slows disease progression in R6/2 Huntington's disease mice. *Annals of Neurology*, 51, 235-242.

Hucklebridge, F.H., Reid, A., Benton, D. and Brain, P.F. (1976). A comparison of the levels of the adrenal catecholamines and aggession in isolated, dominant and subordinate mice. *IRCS Medical Science*, 4, 154.

Hutchinson, E., Avery, A. and VandeWoude, S. (2005). Environmental enrichment for laboratory rodents. *ILAR journal*, 46, 2, 149-161.

Jennings, M., Batchelor, G.R., Brain, P.F., Dick, A., Elliott, H., Francis, R., Hubrecht, R.C., Hurst, J.L., Morton, D.B., Peters, A.G., Raymond, R., Sales, G.D., Sherwin, C.M. and West, C. (1998). Refining rodent husbandry: the mouse. Report of the Rodent Refinement Working Party. *Laboratory Animals*, 32, 233-259.

Korhonen, H., Niemela, P. and Jauhianinen, L. (2001). Effect of space and Floor material on the behaviour of farmed blue foxes. *Canadian Journal of Animal Science*, 81, 189-197.

Latham, N. and Mason, G. (2004). From house mouse to mouse house: the behavioural biology of free living *Mus musculus* and its implications in the laboratory. *Applied Animal Behaviour Science*, 86, 261-289.

Lawrence, A. (1991). Introduction: the biological basis of animals responses to handling. In: *Practical Animal Handling* (Anderson RS, Edney A, eds). Oxford: Pergamon Press, pp 1-13.

Lawrence, E. (2005). Henderson's Dictionary of Biology. Thirteenth Edition, Pearson Education Limited.

Leach, M.C., Ambrose, N., Bowell, V.J. and Morton, D.B. (2000). The development of a new form of mouse cage enrichment. *Journal of Applied Animal Welfare Science*, 3, 81-91.

Lewejohann, L., Reinhard, C., Schrewe, A., Brandewiede, J., Haemisch, A., Görtz, N., Schachner, M. and Sachser, N. (2006). Environmental bias? Effects of housing conditions, laboratory environment and experimenter on behavioural tests. *Genes, Brain and Behaviour*, 5, 64–72.

Martin, P. and Bateson, P. (2007). Measuring behaviour. An introductory guide. *Cambridge University Press*.

Mason, G. (1991). Stereotypies and suffering. *Behavioural Processes*, 25, 103–115.

Mason, G. and Latham, N. (2004). Can't stop, won't stop: is stereotypy a reliable animal welfare indicator? *Animal welfare*, 13, 57-69.

Mason, G., Clubb, R., Latham, N. and Vickery, S. (2007). Why and how should we use environmental enrichment to tackle stereotypic behaviour? *Applied Animal Behaviour Science*, 102, 163–188.

McKernan, R.M., Rosahl, T.W., Reynolds, D.S., Sur, C., Wafford, K.A., Atack, J.R., Farrar, S., Myers, J., Cook, G., Ferris, P., Garrett, L., Bristow, L., Marshall, G., Macaulay, A., Brown, N., Howell, O., Moore, K.W., Carling, R.W., Street, L.J., Castro, J.L., Ragan, C.I., Dawson, G.R. and Whiting, P.J. (2000). Sedative but not anxiolytic properties of benzodiazepines are mediated by the GABA (A) receptor alpha1 Subtype. *Nature NeurosciENCE*, 3, 587-592.

Mench, J.A. (1998). Environmental enrichment and the importance of exploratory behaviour. In: Shepherdson DJ, Mellen JD, Hutchins M, eds. Second Nature: environmental Enrichment for Captive Animals. Washington: Smithsonian Institution Press. p 30-46.

Mering, E., Kaliste-Korhonen, E. and Nevalainen, T. (2001). Estimates of appropriate number of rats: interaction with housing environment. *Laboratory Animals*, 35, 80–90.

Mills, D.S. (2003). Medical paradigms for the study of problem behaviour: A critical review. *Applied Animal Behaviour Science*, 81, 265-277.

Mills, D.S. and Davenport, K. (2002). The effect of a neighbouring conspecific versus the use of a mirror for the control of stereotypic weaving behaviour in the stabled horse. *Animal Science*, 74, 95-101.

Monleon, S., D'Aquila, P., Parra, A., Simon, V.M., Brain, P.F. and Willner, P. (1995). Attenuation of sucrose consumption in mice by chronic mild stress and its restoration by imipramine. *Psychopharmacology*, 117, 453-457.

Nevison, C.M., Hurst, J.L. and Barnard, C.J. (1999). Strainspecific effects of cage enrichment in male laboratory mice (*Mus musculus*). *Animal Welfare*, 8, 361-379.

Newberry, R.C. (1995). Environmental enrichment: Increasing the biological relevance of captive environment. *Applied Animal Behaviour Science*, 44, 229-243.

Öbrink, K.J. & Rehbinder, C. (2000). Animal definition: a necessity for the validity of animal experiments. *Laboratory Animals*, 34, 121–130.

Olsson, A. and Dahlborn, K. (2002). Improving housing conditions for laboratory mice: A review of "environmental enrichment." *Laboratory Animals*, 36, 243-270.

Olsson, A., Nevison, C.M., Patterson-Kane, E.G., Sherwin, C.M., Van de Weerd, H.A. and Würbel, H. (2003). Understanding behaviour: The relevance of ethological approaches in laboratory animal science. *Journal of Applied Animal Behaviour Science*, 81, 245-264.

Paylor, R. (2009). Questioning standardization in science. *Nature methods*, 6, 4, 253-254.

Poole, T. (1997). Happy animals make good science. *Laboratory Animals*, 31, 116-124.

Poole. T. (1998). Meeting a mammal's psychological needs: Basic principles. In: Shepherdson DJ, Mellen JD, Hutchins M, eds. Second Nature: Environmental Enrichment for Captive Animals. Washington: Smithsonian Institution Press. p 94.

Richter, S.H., Garner, J.P. and Würbel, H. (2009). Environmental standardization: cure or cause of poor reproducibility in animals experiments? *Nature methods*, 6, 4, 257-261.

Richter, S.H., Garner, J.P., Auer, C., Kunert, J. and Würbel, H. (2010). Systematic variation improves reproducibility of animals experiments. *Nature methods*, 7, 3, 167-168.

Rochlitz, I. (2005). A review of the housing requirements of domestic cats (*Felis silevtris catus*) kept in the home. *Applied Animal Behaviour*, 93, 97-109.

Rudolph, U., Crestani, F., Benke, D., Brunig, I., Benson, J.A., Fritschy, J.M., Martin, J.R., Bluethmann, H. & Möhler, H. (1999) Benzodiazepine actions mediated by specific gamma-aminobutyric acid (A) receptor subtypes. *Nature*, 401, 796-800.

Russel, W.M.S. and Burch, R.L. (1959). The principles of humane experimental technique. London: Methuen.

Sherwin, C.M. (2002). Comfortable quarters for mice in research institutions. In: Reinhardt V, Reinhardt A, eds. Comfortable Quarters for Laboratory Animals. 9th ed. Washington: Animal Welfare Institute. p 6-17.

Sherwin, C.M. (2004). The influences of standard laboratory cages on rodents and the validity of research data. *Animal welfare*, 13, 9-15.

Stauffacher, M. (1995). Environmental enrichment, fact and fiction. *Scandinavian Journal of Laboratory Animal Science*, 22, 39-42.

Stauffacher, M. (1997a). Housing requirements: What ethology can tell us. In: Van Zutphen LFM, Balls M, eds. Animal Alternatives, Welfare and Ethics. Amsterdam: Elsevier Science BV. p 179-186.

Stauffacher, M. (1997b). Comparative studies on housing conditions. In: O'Donoghue PN, ed. Harmonization of Laboratory Animal Husbandry. London: Royal Society of Medicine Press. p 5-9.

Tsai, P.P., Pachowsky, U., Stelzer, H.D. and Hackbarth, H. (2002). Impact of environmental enrichment in Mice: Effect of housing conditions on body weight, organ weights and haematology in different strain. *Laboratory Animals*, 36, 411–19.

Tsai, P.P., Stelzer, H.D., Hedrich, H.J. and Hackbarth, H. (2003). Are the effects of different enrichment designs on physiology and behaviour of DBA/2 mice consistent? *Laboratory Animals*, 37, 314-327.

Van de Weerd, H.A. and Baumans, V. (1995). Environmental enrichment in rodents. In: Environmental Enrichment Information Resources for Laboratory Animals. AWIC Resource Series 2:145-149.

Van de Weerd, H.A., van Loo, P.L.P., van Zutphen, L.F.M., Koolhaas, J.M. and Baumans, V. (1997a). Preferences for nesting material as environmental enrichment for laboratory mice. *Laboratory Animals*, 31, 133-143.

Van de Weerd, H.A., van Loo, P.L.P., van Zutphen, L.F.M., Koolhaas, J.M. and Baumans, V. (1997b). Nesting material as environmental enrichment has no adverse effects on behaviour and physiology of laboratory mice. *Physiology and Behaviour*, 62, 1019-1028.

Van de Weerd, H.A., Van Loo, P.L.P., Van Zutphen, L.F.M., Koolhaas, J.M. and Baumans, V. (1998). Strength of preference for nesting material as environmental enrichment in laboratory mice. *Applied Animal Behavioural Science*, 55, 369-382.

Van de Weerd, H.A., Aarsen, E.L., Mulder, A., Kruitwagen, C.L.J.J., Hendriksen, C.F.M. and Baumans, V. (2002). Effects of environmental enrichment for mice: Variation in experimental results. *Journal of Applied Animal Welfare Science*, 5, 87-109.

van der Staay, F.J. (2006). Animal models of behavioural dysfunctions: basic concepts and classifications, and an evaluation strategy. *Brain Research Reviews*, 52, 131-159.

van der Staay, F.J. & Steckler, T. (2001). Behavioural phenotyping of mouse mutants. *Behavioural Brain Research*, 125, 3–12.

van der Staay, F.J. and Steckler, T. (2002). The fallacy of behavioural phenotyping without standardisation. *Genes, Brain and Behaviour*, 1, 9–13.

Van der Staay, F.J., Arndt, S.S. and Nordquist, R. (2009). Evaluation of animal models of neurobehavioural disorders. *Behavioural and Brain functions*, 5, 11.

van der Staay, F.J., Arndt, S.S. and Nordquist, R.E. (2010). The standardization-generalization dilemma: a way out. *Genes, brain and behaviour*, accepted article.

Van Loo, P.L.P., Kruitwagen, C.L.J.J., Koolhaas, J.M., Van de Weerd, H.A., Van Zutphen, L.F.M. and Baumans, V. (2002). Influence of cage enrichment on aggressive behaviour and physiological parameters in male mice. *Journal of Applied Animal Behavioural Science*, 76, 65-81.

Van Zutphen, L.F.M. (2003). Inleiding. In: van Zutphen, L.F.M., Baumans, V. and Beynen, A.C., eds. Handboek proefdierkunde. Proefdieren, dierproeven, alternatieven en ethiek. 4th edition, Elsevier gezondheidszorg.

Vestergaard, K.S., Skadhauge, E. and Lawson, L.G. (1997). The stress of not being able to perform dustbathing in laying hens. *Physiology and Behaviour*, 62, 413-419.

Vickery, S.S. and Mason, G.J. (2003). Behavioural persistence in captive bears: implication for reintroduction. *Ursus*, 14, 35-43.

Wahlsten, D. (2001). Standardizing tests of mouse behaviour: reasons, recommendations, and reality. *Physiology and Behaviour*, 73, 695–705.

Wahlsten, D., Metten, P., Phillips, T.J., Boehm, S.L., Burkhart-Kasch, S., Dorow, J., Doerksen, S., Downing, C., Fogarty, J., Rodd-Henricks, K., Hen, R., McKinnon, C.S., Merrill, C.M., Nolte, C., Schalomon, M., Schlumbohm, J.P., Sibert, J.R., Wenger, C.D., Dudek, B.C. and Crabbe, J.C. (2003). Different data from different labs: lessons from studies of gene-environment interaction. *Journal of Neurobiology*, 54, 283–311.

Wahlsten, D., Bachmanov, A., Finn, D.A. and Crabbe, J.C. (2006). Stability of inbred mouse strain differences in behaviour and brain size between laboratories and across decades. *Proceedings in Natl. Academic Science USA*, 103, 44, 16364–16369.

Weed, J.L. and Raber, J.M. (2005). Balancing animal research with animal wel-being: establishment of goals and harmonization of approaches. *ILAR journal*, 46,2, 118-128.

Wiepkema, P.R. and Koolhaas, J.M. (1993). Stress and animal welfare. *Animal Welfare*, 2, 195-218.

Wolfer, D.P., Litvin, O., Morf, S., Nitsch, R.M., Lipp, H.P. and Würbel, H. (2004). Laboratory animal welfare: cage enrichment and mouse behaviour. *Natue*, 432, 821-822.

Würbel, H. (2000). Behaviour and the standardization fallacy. *Nature Genetics*, 26, 263.

Würbel, H. (2001). Ideal homes? Housing effects on rodent brain and behaviour. *Trends in Neuroscience*, 24, 207-211.

Würbel, H. (2002). Behavioural phenotyping enhanced - beyond (environmental) standardization. *Genes, Brain and Behaviour*, 1, 3–8.

Würbel, H., Chapman, R. and Rutland, C. (1998). Effect of feed and environmental enrichment on development of stereotypic wire-gnawing in laboratory mice. *Applied Animal Behaviour Science*, 60, 69-81.

Würbel, H. and Garner, J.P. (2007). Refinement of rodent research through environmental enrichment and systematic randomization. Available: http://www.nc3rs.org.uk/news.asp?id=395.

Young, R.J. (2003). Environmental enrichment for captive animals. UFAW Animal Welfare Series. London: Blackwell Science Ltd.