INHIBITION

AND

LEARNING

Edited by

R. A. BOAKES and M. S. HALLIDAY

Laboratory of Experimental Psychology
University of Sussex,
Sussex, England

1972

ACADEMIC PRESS LONDON NEW YORK

Some Ideas Concerning Physiological Mechanisms of so-called Internal Inhibition

in: R. A. Boakes and M. S. Halliday (ed.),
Inhibition and learning. Acad. Press, London, p. 341-357

1 3

J. Konorski

Department of Neurophysiology, Nencki Institute of Experimental Biology, Warsaw, Poland

I. Introduction

When discussing the physiological mechanisms of inhibitory processes involved in learning we should begin with a short survey of the occurrence of these processes in more fundamental functions of the nervous system observed mostly in acute experiments performed in anaesthetized or immobilized animals. For these experiments give us indubitable evidence that inhibitory processes are no less, and may be even more, ubiquitous than excitatory processes, and that the normal activity of the nervous system is thoroughly impregnated by their presence.

If we try to categorize the conditions in which inhibitory processes occur, we may specify at least the four following cases.

1. Reciprocal inhibition. There is a great body of evidence to show that most nervous "centres" (i.e. sets of neurons endowed with the same functions) are arranged in antagonistic pairs, such that excitation of one centre inhibits the other one and vice versa. The first example of this general rule was provided by Sherrington (1947) in spinal reflexes with the help of crude stimulus-response techniques; methods of detecting this type of inhibition have since become more and more refined and include the method of recording hyperpolarization in nerve cells by intracellular microelectrodes (Eccles, 1964). Typical examples of this reciprocal inhibition at various levels of the nervous axis are: flexion versus extension of the limbs,

inspiration versus expiration, hunger versus satiation, reciprocal relations of centres controlling bodily temperature, and last but not least, arousal versus somnolence. This type of inhibition is characterized by the reciprocity of the mutual relations between a pair of centres, which Sherrington used to call "subcentres".

2. Antagonisms between centres (or rather functional systems) which are not specifically paired. Again the simplest Sherringtonian model of this relation is inhibition of the scratch reflex, produced by a nociceptive stimulus. To turn to a higher level of nervous integration, we can indicate the inhibitory influence of fear reflexes upon *any* other drive reflexes (hunger, anger, somnolence, sexual drive).

3. Unidirectional inhibition. There are structures in the nervous system which exert overall inhibitory effects upon other structures. In the majority of cases we deal here with the inhibitory influence of higher structures, including the cerebral cortex, upon lower structures. Disinhibition of the function of these lower centres by the removal of the higher centres is called the "release phenomenon". Decerebrate rigidity and sham rage are typical examples of this release.

4. Lateral inhibition. This is probably the most ubiquitous type of inhibitory process within the nervous system. It is produced by a special type of inhibitory neurons with short and widely ramifying axons, exerting inhibitory influence upon neighbouring long-axon neurons. Lateral inhibition accounts for delicate motor adjustments both in the spinal cord and the motor cortex, as well as for the sharpening of contrast in afferent functions. Konorski (1967) emphasized the important role of lateral inhibition in perceptual processes at the highest levels of the nervous hierarchy.

Keeping in mind these four (at the minimum) categories of inhibitory process, we turn now to the discussion of inhibitory processes as they are manifested in complex forms of behaviour in higher animals.

There is no doubt that Pavlov was virtually the first scientist who utilized the concept of inhibition in behavioural experimentation by introducing his notions of external and internal inhibition. By internal inhibition he meant those inhibitory processes which are not ready-made as a result of ontogenetic development (as is the case with external inhibition), but which are learnt whenever the external conditions require the animal to suppress some of its inborn or acquired behavioural acts. According to various inhibitory training procedures he distinguished extinction, differentiation (including so-called conditioned inhibition), and inhibition of delay.

I shall not dwell upon the original Pavlovian theory of internal inhibition, an evaluation of which was presented in my monograph published in the forties (Konorski, 1948). In that monograph I put forward a concept of the mechanism of internal inhibition which assumed that it consists in the

formation of inhibitory synaptic connections between the "centre" of the conditioned stimulus (CS) and the centre of the unconditioned stimulus (US). An important assumption, based on relevant experimental evidence, was that the formation of an inhibitory conditioned reflex (CR) does not consist in the *transformation* of excitatory synaptic paths between the two centres into inhibitory ones, but rather in the *addition* of inhibitory connections to the previously developed excitatory connections. To put it in a different way, we may say that the "inhibitory CR" established by the procedure of extinction, differentiation or inhibition of delay is, as a matter of fact, a mixed excitatory-inhibitory CR, its reflex-arc consisting of both types of synaptic connection between the CS centre and the US centre.

This assumption accounts for the gradual elaboration of the inhibitory CR and its "disinhibition" (temporary or permanent) whenever the newly established inhibitory connections are for some reason outweighed by previously established excitatory connections. My previous monograph (Konorski, 1948) was mainly concerned with demonstrating that the great bulk of data on internal inhibition gathered in Pavlov's laboratories could be satisfactorily accounted for by this concept. Thus it was postulated that both excitatory and inhibitory synaptic connections are established either in the course of phylogenetic development, or in the individual life of the organism as a result of his experiences.

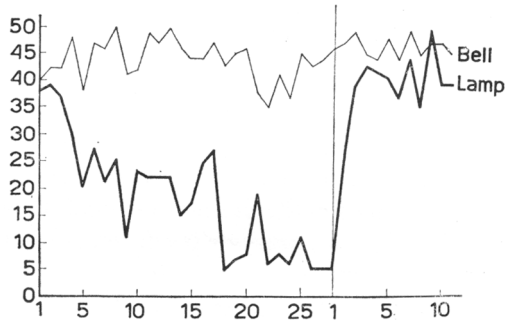
My previous concept was developed about 25 years ago, before the postwar work of our laboratory started in the Nencki Institute. It now appears to be inadequate and has been replaced by a quite different concept, which seems to me both more adequate for the explanation of numerous experimental facts gathered in the meantime, and more reasonable from the physiological point of view. This new concept has been described in detail (Konorski, 1967), and since I have not so far changed my views upon this subject, the present considerations will be based on the same ideas and experimental evidence which were discussed in that book.

II. Experimental Data on the Transformations of Excitatory into Inhibitory CRs, and Vice Versa

When testing my previous hypothesis concerning the mechanism of internal inhibition, Konorski and Szwejkowska (1950) came across the following facts:

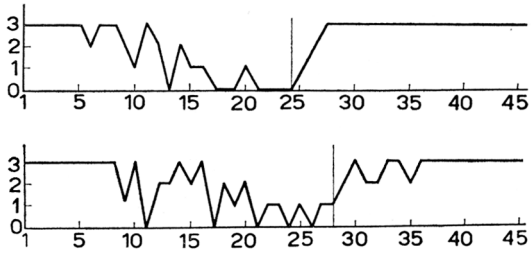
In experiments on dogs in a standard CR chamber a number of classical food CRs were established and measured by the magnitude of salivary response. When the CRs became stable, one of the CSs was presented once or twice per session, without reinforcement, while other CSs continued to be reinforced. This chronic extinction lasted for about one month until a definite extinction level was reached. Thereupon the excitatory CR to

that CS was restored by reinforcing it again in similar conditions. It was found that while resistance to extinction of the excitatory CR was considerable, resistance to its restoration was minimal, since after one or two reinforced trials the previous level of the salivary response was attained (Fig. 1).



Fio. 1. An example of chronic extinction and restoration of classical food CR. Abscissa: experimental sessions with one extinction trial per session. Ordinate: conditioned salivation in arbitrary units. The vertical line denotes the beginning of restoration of the CR to lamp. Lamp is the extinguished CS; Bell is the control CS, which always just precedes an extinction trial. Note that the process of extinction is slow and irregular, while restoration of the extinguished CR is immediate (from Konorski, 1967).

The same asymmetry of the rate of extinction and restoration was obtained in experiments by Konorski and Szwejkowska (1952a) with defensive classical CRs, the unavoidable US being a shock delivered to the paw (Fig. 2).



Fm. 2. Chronic extinction and restoration of classical defensive CR. Abscissa: experimental sessions. Ordinate: the number of positive responses to the extinguished and restored CS. In each session this CS was presented three times among positive CSs. The vertical line denotes the beginning of restoration of the CR. Each graph denotes the experimental results from one dog. Note the slow and irregular course of extinction and much more rapid course of restoration of the CR (from Konorski and Szwejkowska, 1952a).

In another series of experiments a bell (S1) was a positive (excitatory) CS signalling food. When the CR to this CS was firmly established, a bell

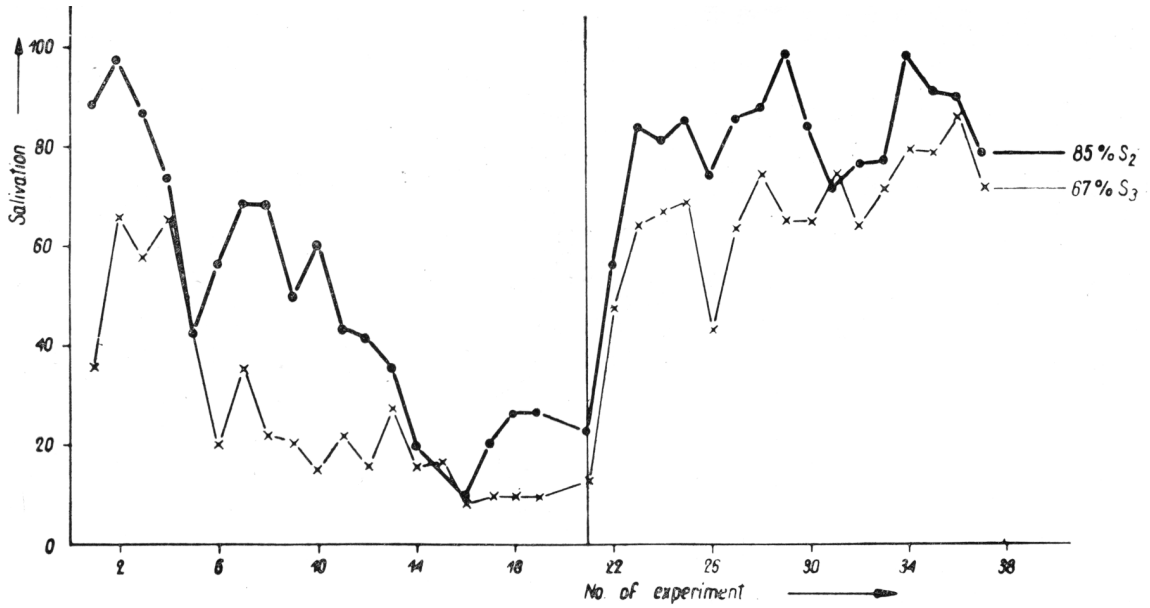


FIG. 3. Extinction of two stimuli similar to a positive CS and their transformation into positive CSs. Abscissa: experimental sessions. Ordinate: salivation as a percentage of the effect of the positive CS (S_1). The vertical line denotes the beginning of positive conditioning. Note the stronger resistance to extinction of a CR to a stimulus more similar to the positive CS, S2, than to a stimulus less similar, S_3, and the weaker resistance to conditioning of S_2 than to conditioning of S_3 (from Szwejkowska, 1959).

of different sound (S_2) and a buzzer (S_3) were presented without reinforcement among other positive CSs. As seen in Fig. 3, S_3 was less similar to S~ than ~2, as judged by weaker generalization and weaker resistance to extinction. When both S_2 and S_3 were converted into positive CSs by reinforcing them, it appeared that resistance to conditioning was much weaker for ~2 than for 53, that is the closer the differentiated CS to the original CS, the easier its transformation into the positive CS (Szwejkowska 1959).

Finally, in the third series of experiments, after the formation of an excitatory food CR to a given stimulus, a stimulus quite different from the original CS (as judged by the lack of generalization) was introduced and repeatedly presented without reinforcement. When, after this nonreinforced training, the stimulus was converted into an excitatory CS by food reinforcement its resistance to conditioning was extremely strong, and this stimulus practically never produced a strong and regular conditioned response (Fig. 4). It should be noted that in the presence of that stimulus the animals refused to take food, waiting till the stimulus was discontinued.

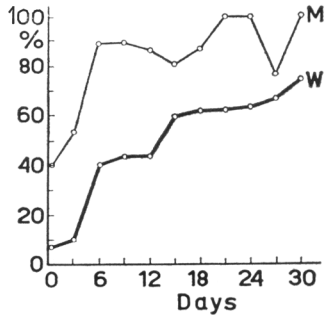


Fig. 4. The formation of classical food CR to a new stimulus (Metronome) and to a stimulus repeatedly presented without food reinforcement (Whistle). Abscissa: experimental sessions. Ordinate: conditioned salivation as a percentage of that of a well-established CR. Note that a new stimulus (M) elicits from the very beginning considerable salivation (pseudo-conditioning), rapidly attaining the level of the control CR, whereas the nonreinforced stimulus (W) originally elicits a negligible salivary response which very slowly increases when the stimulus is reinforced. The irregularity of the responses is not seen, because each point denotes the average of three sessions (from Konorski, 1967).

Experimental neuroses developed occasionally under these conditions (Konorski and Szwejkowska, 1952b).

An important conclusion drawn from these experiments is that a stimulus presented without reinforcement among other stimuli which are duly reinforced, and being outside the field of their generalization, is not a "neutral"

stimulus (as was thought before), but acquires strong inhibitory properties, as judged from its resistance to conditioning. Thus, while a stimulus which had previously signalled the US, and then stopped doing so because of nonreinforcement, easily regains its signalling capacity, a stimulus which was presented among positive CSs but always signalled the lack of the US, could not thereafter be converted into a regular and reliable signal heralding the occurrence of the US.

According to these data the original version of the concept of inhibitory conditioning should be modified by distinguishing secondary inhibitory CRs which arise when excitatory CSs cease to be reinforced by the US, and primary inhibitory CRs, which arise when the corresponding CSs are never reinforced. Although the primary and the secondary CSs may be phenomenologically indistinguishable, because both of them produce no salivary response, they *can* be distinguished when we convert them into excitatory CSs. While the secondary inhibitory CS is very quickly transformed into the excitatory CS, the primary inhibitory CS is resistant to such a transformation.

III. A New Theory of Internal Inhibition

We might be satisfied with this improved concept of internal inhibition when dealing with transformations of *homogeneous* CRs, either alimentary, or defensive. The situation is, however, changed, when we deal with transformations of *heterogeneous* CRs, for instance if we transform alimentary CRs into defensive CRs and vice versa (Konorski and Szwejkowska, 1956; Konorski, 1967). It has been found that the same rules stated above for excitatory and inhibitory CRs hold true with regard to food-excitatory and shock-excitatory CRs. In fact, whereas the *formation* of either alimentary or defensive CRs to new stimuli is generally rapid, the *transformation* of CSs from one of these categories into the other one encounters great resistance and the new CR is never fully attained. Moreover, it has been found that a transformed CS has a mixed nature, being both alimentary and defensive, and this or that aspect of this CS may become manifest, depending on whether it is presented against an alimentary background or against a defensive background.

This being so, one is tempted to propose a theory which would account in the same manner for both kinds of transformation, namely excitatoryinhibitory transformations in homogeneous CRs and alimentary-defensive transformations in heterogeneous CRs. Such a theory can indeed be established if we admit that non-reinforcement of a given stimulus, in a situation in which other stimuli *are* reinforced by a definite US, means that this stimulus is reinforced by the *lack of the US* (denoted as no-US). In physiological terms it maybe assumed that there are two reciprocally related centres

(for instance, food and no-food centres), the first one being activated by the taste of food and the other by no-food in the mouth. The "excitatory" CS is a CS the centre for which is connected with the US centre, whereas the "inhibitory" CS is a CS the centre for which is connected with the no-US centre. This theory has been recently developed (Konorski, 1967), and it has been found that it accounts much better for the available experimental data than does the previous theory, which postulated the formation of both excitatory and inhibitory connections between the CS centre and a unique US centre.

In order to show the advantages of the new concept over the previous one, let us consider in more detail the processes of the formation of excitatory and inhibitory CRs and their mutual transformations on the basis of the experimental data described above. As explained above, the formation of a primary excitatory CR carried out by pairing the CS with the US is generally a rapid process requiring a small number of trials. Extinction of the CR, on the contrary, is a rather lengthy process requiring several dozen non-reinforced trials (cf. Figs. 1 and 2). If extinction is supposed to be due to the formation of connections between the CS centre and the no-US centre, this slowness is understandable, as a result of the antagonistic relationship between the US and the no-US centres. In fact, in the first stage of extinction the CS strongly activates the US centre, which in turn strongly inhibits the no-US centre. Accordingly, although the pairing of the CS and the no-US does occur, the connections between the two centres cannot be formed because of the nonreactive state of the no-US centre. This situation, however, cannot last indefinitely. After all, the US does *not* follow the CS and accordingly the activation of the no-US centre begins to take place, first after the cessation of the CS and then in the presence of the CS itself. This activation, in turn, inhibits the US centre and thus more and more gives an upper hand to the no-US centre and allows connections between the CS centre and the no-US centre to be formed. Finally, these connections become so abundant that the US centre becomes completely inhibited during the action of the CS.

What are the experimental consequences of the proposed mechanism in comparison with the hypothesis assuming the formation of inhibitory connections between the CS centre and the US centre, as postulated by the previous theory?

First, the theory assuming the formation of inhibitory connections between the CS and US centres would predict a regular and rather linear decrease of the magnitude of the conditioned response, since each nonreinforced trial adds a "quantum" of inhibition to the CS—US connections. On the contrary, the theory of two reciprocal centres being involved in extinction predicts a quite different course of events: the process pf extinc-

tion should be very slow in the first extinction trials and then should gain momentum in later trials, thus being completely non-linear.

Secondly, in the middle of the process of extinction, when there is a balance between the CS—US connections and the CS-no-US connections, the process *must* be utterly irregular because even a small dominance in the activation of one of the two centres would immediately give it the upper hand, because of the reciprocal inhibition of the other centre. Thus oscillations in the magnitude of conditioned responses should in the transition stage, be much above the reasonable chance level, because each oscillation is amplified by positive feedback due to the instability of the whole system.

Thirdly, if the duration of the CS on each trial lasts a dozen seconds or more, as is the case in most experiments with salivary CRs, then the irregularity of conditioned responses should be observed not only between successive trials but also within trials. For instance, at the beginning of its action the CS might produce a full-sized salivary response, which may stop abruptly after a few seconds when the inhibitory process starts to dominate the excitatory process, or, more rarely, the sequence of the processes may be the reverse.

My own experience in CR experimentation, acquired before the present concept had been developed, is in full agreement with these predictions and in full conflict with the previous theory. Like all learning theorists who tried arduously to construct "learning curves" by smoothing down all the irregularities of learning processes, I had also believed in the real existence of these curves and that they reflected the "true" course of learning. Now I think, on the contrary, that these curves falsify reality by concealing the irregularity of the learning process, an irregularity which is inherent in its very nature.

Now, what is the situation when the extinguished CR is restored by reinforcement of the CS? According to our previous view this restoration should work in the opposite direction to extinction, adding an excitatory quantum in each successive trial. In consequence the process should again be linear and perhaps symmetrical to the process of extinction, if excitatory quanta and inhibitory quanta are of equal value. On the other hand, according to our present view, the situation is as follows: the CS centre is now connected with the US centre and the no-US centre, and therefore the dominance of one of those connections over the other one depends on a relatively small difference. The system, therefore, becomes bistable with a relatively strong preponderance of one of the two states depending on minor factors.

We are confronted with a quite different situation when dealing with a primary inhibitory CS, namely a stimulus which is presented among excitatory CSs reinforced by a given US, but is never followed by this US.

In that case the non-reinforced CS becomes a consistent signal for the no-US, which means that its centre forms connections with the no-US centre only. Again, this CS does not differ very much from an extinguished CS, unless it is converted into the excitatory CS. Then we may notice that its resistance to conditioning is exceedingly strong and it hardly ever acquires the stability and reliability of a primary excitatory CS. The reason for this resistance to conditioning lies in the fact that the activation of the no-US centre produced by this CS strongly inhibits the US centre and thus prevents the formation of connections between the CS centre and the US centre. When the excitatory CR to the former inhibitory CS is eventually established it is, as a rule, irregular because whenever for some reason the US centre is not activated this immediately leads to the activation of the no-US centre which further inhibits the US centre.

The assumption that, apart from the units activated by actual stimuli of various modalities, there are units which are thrown into action when these stimuli are discontinued, or even not operating at all for a length of time, may seem paradoxical, but only at first glance. As a matter of fact, there exist numerous units in the perceptual areas of the brain which discharge "spontaneously", in spite of the absence of any observable actual stimuli, but are immediately silenced, when a given stimulus is presented. When this stimulus is discontinued, the unit resumes its activity with an increased rate (the so-called "off-effect"), returning thereafter to its "spontaneous" moderate rate of firing. In other words, such an "off-unit" behaves exactly in the same way as an "on-unit", except that it does not react to an actual stimulus but to its absence. These off-units obviously play a most important role in perceptual processes, because they actively announce that "nothing happens" in the given perceptual field, information which may be as valuable as that something does happen. Accordingly, when we claim that along with "centres" (that is groups of units) representing tastes of particular kinds of food, there are "centres" which represent the absence of food in the mouth, we stick to a general principle of the activity of the nervous system. We should add only that the absence of food in the mouth activates the off-taste units only when the subject turns its attention to its gustatory perceptual field, which normally occurs when it is hungry. This is why one is aware of the emptiness of one's mouth when one is looking for food.

The hypothesis proposed above is in good agreement with the fact that, as found in many experiments on various species of animals (cf. Brutkowski, 1965), ablation of a specific area within the prefrontal cortex leads to a dramatic impairment of inhibitory alimentary CRs. Our explanation of this phenomenon is as follows (Konorski, 1967, 1971): it is assumed that the orbital part of the prefrontal area in monkeys and the medial part of this area in dogs is an extension of the limbic system. Its role is the higher con-

trol of alimentary behaviour and, in particular, inhibition of conditioned alimentary responses, in spite of humoural hunger, in those situations in which food is completely unavailable. In other words this area becomes connected with centres of all those external stimuli which signal the unavailability of food, and its activation exerts an inhibitory influence upon humoural hunger centres. Accordingly, if this area is removed, the animal is

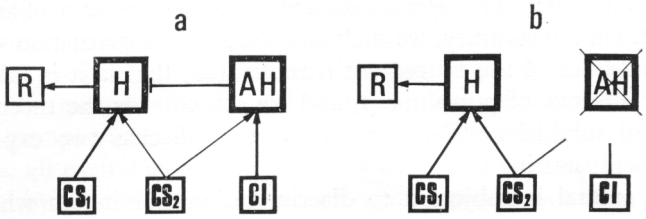


FIG. 5. Block model of the mechanism of inhibitory instrumental CR (a) and its impairment after prefrontal lesion (b). CS_1, excitatory CS centre; CS_2, inhibitory CS centre; CI, conditioned inhibition centre; H, hunger system; AH, antihunger centre situated in the prefrontal extension of the limbic system; R, instrumental response centre. Arrows, excitatory connections; stopped lines, inhibitory connections. Thin lines denote weak connections.

unable to suppress hunger by no-food CSs, and this deficit is manifested under experimental conditions by disinhibition of inhibitory CRs. The model of this action is shown in Fig. 5.

IV. The Problem of Motor-act Inhibition

The important conclusion which follows from our considerations is that plasticity of the brain consists in the formation of only excitatory synapses, while inhibitory CRs *utilize* inborn inhibitory connections which link either reciprocally related centres or higher inhibitory centres with subordinate ones. Of course, we cannot generalize this thesis until we can prove that there are no instances of learning based on the formation of new inhibitory synaptic contacts.

So far we have been concerned only with inhibitory effects exerted upon US centres and drive centres. However, animal and human behaviour consists mainly of purposeful (voluntary) movements, represented in a simplified model by instrumental conditioned responses. Shaping this behaviour includes not only learning *to perform* particular motor acts in response to particular situations, but also *not* to perform these acts in the presence of "prohibitions" commanding: don't do this or that, because if you don't you may get a reward, but if you do you will be punished (either on earth or in hell). Thus we have learnt during life to restrain many motor activities, and the problem arises whether all these restraints are established

by the formation of direct inhibitory synapses on the neuronal groups controlling these activities, or whether excitatory synapses are formed on the neuronal groups controlling antagonistic activities.

To avoid misunderstanding, the present discussion is not concerned with the inhibition of instrumental responses when the animal stops performing a movement which is not rewarded. Here the performance of the movement is discontinued because the *drive* CR is extinguished, that is, there is no motivation which is the *spiritus movens* of the performance of any instrumental act. On the contrary, we shall now deal with the situation where the non-performance of the movement *is* motivated, that is it is followed by reward or absence of punishment, and we ask what is the mechanism of *that* sort of inhibition of motor acts. Let us discuss two experimental procedures illustrating this situation.

1. The animal is subjected to discrimination training in which he is required, in order to receive food, to perform a given movement (say, leg flexion) in response to one stimulus, while in the presence of the other stimulus he is obliged not to perform that movement in order to receive food. This training leads to the animal actively refraining from the performance of that movement by executing the antagonistic movement (extension). We have good evidence to show that in this type of procedure a different mechanism is in operation from that involved in the Pavlovian type of differentiation (when the negative CS is simply not reinforced), because the symmetrically reinforced differentiation task is impaired after quite different prefrontal lesions (Dabrowska, 1971).

2. The dog is trained in classical alimentary conditioning to a certain CS. Now, from time to time, in the presence of that CS we passively lift the dog's left foreleg with the aid of some apparatus, and if the leg is lifted food is not offered. Very soon we notice that the animal keeps his left foreleg immobile throughout the experimental session; at the onset of the CS the leg is extended and the extension grows stronger when the leg is pulled upwards. It is clear that the animal resists the bending of his foreleg with his whole strength (Konorski and Miller, 1936).

The situation roughly the same when the animal refrains from performing a certain movement because this movement is followed by a noxious stimulus. Here too we observe the performance of the antagonistic movement which increases in strength whenever the animal is pulled or pushed to perform the danger-provoking motor act.

As we see, in all these situations we have clear evidence that the muscles antagonistic to those participating in the performed motor act are brought into action and that this is the way by which this motor act is inhibited. Therefore, again we have no evidence to show that new inhibitory synaptic connections are formed between the CS centre and the centre of the inhi-

bited motor act. This finding suggests that perhaps in *all* cases of motor-act inhibition the mechanism is the same and this inhibition occurs by the mediation of excitation of groups of neurons antagonistic to those eliciting that motor act.

V. Conclusion

If we try to explain in physiological terms the class of phenomena referred to as inhibitory CRs on the basis of a connectionist theory of the functioning of the nervous system, we are confronted with the following problem.

We have good reason to believe that conditioning and learning are based on the increase of transmissibility of synaptic contacts between particular groups of neurons, regardless of the detailed nature of this process. Thus, all excitatory CRs and, more generally, all associations can be understood by reference to this concept, again regardless of what groups of neurons take part in particular learning processes.

On the other hand, in experiments on CRs (including habit formation) we encounter a sort of "negative learning" when a subject is trained to suppress his response, if it is for some reason maladaptive. To account for this phenomenon we have to choose between two possible mechanisms. One mechanism, in which I believed for many years, is that learning consists in the formation of either excitatory or inhibitory synaptic contacts between the CS centre and the US centre, the former being responsible for excitatory CRs, the latter for inhibitory CRs. My previous monograph (Konorski, 1948) advocated this approach, which I tried to support by the known experimental data. The second possible mechanism is based on the assumption that inhibitory CRs are established by the formation of excitatory connections between the CS centre and the no-US centre, the latter centre being reciprocally related to the former one. In our preceding discussion we have seen that the latter theory accounts much better for experimental data concerning inhibitory CRs than does the previous theory based on the formation of inhibitory synapses between the CS and the US units.

Now, analysing in more detail inhibitory CRs we may find that they can be divided into three somewhat differing groups:

First, we may deal with inhibition in classical consummatory CRs, as represented by food CRs, measured by salivation, or shock CRs measured by flexion of the leg to which shock is delivered. Here, as we said before, the inhibitory CR is established by formation of connections between the central representation of a given CS and that of no-food or no-shock, respectively.

Secondly, we may deal with inhibition within drive CRs, as represented by hunger CRs or fear CRs. As we know, instrumental responses in approach or avoidance training, respectively, are the best overt indicators

of these CRs. If in a given drive situation a given CS is never reinforced by food or by a noxious agent, respectively, then this CS becomes a signal of hunger anti-drive or fear anti-drive, manifested by the absence of instrumental approach responses or avoidance responses. Of particular interest are the anti-drive hunger CRs when a subject is confronted with a situation in which food is completely unavailable, and therefore his hunger drive is inhibitied by the action of a higher order inhibitory centre localized in the prefrontal cortex.

Finally, we may deal with inhibition within instrumental responses themselves, when the performance of a given motor act is maladaptive and therefore should be inhibited. According to the experiments involving such a situation in animals, we have many reasons to believe that inhibition of the maladaptive motor acts occurs owing to excitation of neurons eliciting antagonistic motor acts. We cannot, however, be certain whether this is the only way of opposing the maladaptive behaviour, or whether there exists a direct inhibitory mechanism controlling this behaviour.

VI. Postscript

The above concept concerning the physiological mechanism of internal inhibition seems to be suitable for the explanation of a number of facts presented by other contributors. I shall discuss a few of these facts and try to show how I would explain them within the framework of my concept.

I shall begin with a discussion of Bitterman's data (Chapter 6) concerning the repeated reversal training of the Go; No go differentiation, in which, of the two CSs (for instance, red and green), either one or the other was reinforced by food. It has been found that the first reversal takes longer to train than the following ones. The explanation of this is as follows. In the original training the positive CS centre became connected with the US centre, and the negative CS centre, with the no-US centre. On the other hand, in the first reversal, the so far positive CS centre must have become connected with the no-US centre, while the so far negative CS centre must have become connected with the US centre. As shown in my chapter, this training encounters inherent resistance. However, after the first reversal, and even more after a few successive reversals, the situation changes:
now *each* of the two CSs is connected both with the US centre and the no-US centre, and therefore the slight dominance of one of the respective connections, produced by one or a few appropriate trials is sufficient for the animal's correct response.

The fact that a pigeon develops a "positive presumption" in response to both CSs, tending at the beginning of each reversal to react to both of them, and then eliminating the conditioned response to the now negative CS, is understandable. For, other things being equal, the connections of

both CS centres with the US centre are slightly stronger than those with the no-US centre.

I should now like to turn to the results reported by Wagner and Rescorla (Chapter 12), and discuss their beautiful experiments on conditioned inhibition. In one experiment A is a strong fear CS and B is a weak one; the stimulus X precedes trials in which the shock to A or B is not given. According to our concept X forms "pure" connections with the no-fear (or safety) centre. Wagner and Rescorla show that X becomes a stronger inhibitory CS when it is paired with A than when paired with B. This fact may be explained by reference to our thesis that all types of connections (whether with US or no-US) are better developed against the background of a strong drive, than against the background of a weak drive. Since X is a primary inhibitory CS (and not a secondary inhibitory CS *transformed* from the primary excitatory CS) it is clear that this CS does not form connections with the US centre which would dwarf the connections established with the no-US centre.

Finally, I wish to comment on the results described by Halliday and Boakes (Chapter 3), and by Terrace (Chapter 4), concerning the presence or absence of the "behavioural contrast phenomenon" in discrimination learning depending on different experimental procedures.

In my recent monograph I have explained the contrast phenomenon as due to the fact that occasional lack of reinforcement of a CS increases the hunger drive for a whole experimental session (Ronorski, 1967). Therefore, when we train an animal in Pavlovian differentiation of two similar stimuli CS ± and CS —,the centre of CS —,is connected with the US centre, owing to generalization, and with the no-US centre, owing to nonreinforcement of that stimulus by food. Since CS – was originally an excitatory CS as a result of generalization it still elicits a conditioned drive reflex, which is not inhibited by the consummatory food response, and therefore increases the drive level of the subject. This is why, according to Terrace, "subjects who learn with errors exhibit emotional responses in the presence of S".

On the other hand, when the animal learns the discrimination "without errors", CS— is outside the field of generalization of CS ±, being what I have called a "primary inhibitory CS" (Konorski, 1967). Since such a stimulus was never connected with either food, or hunger drive, its properties are quite different from those of the extinguished CS, because it not only fails to increase the hunger drive, but may even decrease it. This is why the subject "tends to squat down . . . and quietly await the next presentation of S ±" (Terrace).

The situation encountered in the Halliday and Boakes experiments is somewhat different. These authors first trained the pigeons to peck the key to S1, and S_2, and thereafter in the presence of S_2, either withheld

reinforcement (extinction subjects), or offered "free reinforcement" (Free VI subjects). It appears that contrast in S1 trials was observed only when food was withheld, but not when given gratis in the presence of S_2.

How are these facts to be explained? As far as the extinction subjects are concerned, extinction of the instrumental response to S_2 leads, in our view, to an increase of the hunger drive manifested by the increased rate of keypecking in response to S1. With regard to Free VI subjects, the situation is different. Since food is offered gratis in the presence of S_2, a classical CR is established to this stimulus, owing to which the instrumental CR is suppressed but not extinguished (cf. Konorski, 1967). Accordingly, S_2 in no way produces an increase in the hunger drive, since it is accompanied by the consummatory food response.

To sum up, in the above experiments we were confronted with three types of "negative" stimuli (with regard to the instrumental response):

(i) a stimulus which stopped eliciting a response because of the withholding of reinforcement;
(ii) a stimulus which stopped eliciting a response because food was offered gratis;
(iii) a stimulus which never elicited a response because, owing to a special procedure, it was outside the field of generalization of a positive CS.

According to my view, only the first type of stimulus produces an increase in hunger drive when it is presented along with positive CSs. The second and third type do not produce this effect, because, respectively, either drive is satisfied by presentation of food gratis, or the subject does not hope to receive food, since it was never presented in the presence of that stimulus.

References

Brutkowski, S. (1965). Functions of prefrontal cortex in animals. *Physiological Review,* 45, 721—746.

Dabrowska, 1. (1971). Dissociation of impairment after lateral and medial prefrontal lesions in dogs. *Science,* 171, 1037—1038.

Eccles, J. C. (1964). *The Physiology of Synapses.* Berlin, Gottingen Heidelberg: Springer Verlag.

Konorski, J. (1948). *Conditioned Reflexes and Neuron Organizations.* London: Cambridge University Press; second printing New York and London: Hafner Publishing Company, 1968.

Konorski, J. (1967). *Integrative Activity of the Brain.* Chicago and London: The University of Chicago Press; second printing, 1970.

Konorski, 1. (1971). The role of prefrontal control in programming of motor behaviour. In: J. D. Maser (Ed.), *Efferent Organization and Integrative Behaviour.*

Konorski, J. and Miller, S. (1936). Conditioned reflexes of motor analyser. (In press).

Russian with English summary.) *Trudy fiziologitcheskikh laboratoryi akademika J. P. Pavlova,* 6/1, 119—278.

Konorski, J. and Szwejkowska, G. (1950). Chronic extinction and restoration of conditioned reflexes. I. Extinction against the excitatory background. *Acta Biologiae Experimentalis,* 15, 155—170.

Konorski, J. and Szwejkowska, G. (1952a). Chronic extinction and restoration of conditioned reflexes. III. Defensive motor reflexes. *Acta Biologiae Experimentalis,* 16, 91—94.

Konorski, J. and Szwejkowska, G. (1 952b). Chronic extinction and restoration of conditioned reflexes. IV. The dependence of the course of extinction and restoration of conditioned reflexes on the "history" of the conditioned stimulus. *Acta Biologiae Experimentalis,* 16, 95—113.

Konorski, 5. and Szwejkowska, G. (1956). Reciprocal transformations of heterogenous conditioned reflexes. *Acta Biologiae Experimentalis, 17,* 141—165.

Sherrington, C. S. (1947). *The Integrative Action of the Nervous System.* 7th Ed. London: Cambridge University Press.

Szwejkowska, G. (1959). The transformation of differentiated inhibitory stimuli into positive conditioned stimuli. *Acta Biologiae Experimentalis,* 19, 151—159.