A Neural Theory of Punishment and Avoidance, II: Quantitative Theory

STEPHEN GROSSBERG
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Communicated by Richard Bellman

ABSTRACT

Quantitative neural networks are derived from psychological postulates about punishment and avoidance. The classical notion that drive reduction is reinforcing is replaced by a precise physiological alternative akin to Miller's "Go" mechanism and Estes's "amplifier" elements. Cell clusters $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ are introduced which supply negative and positive incentive motivation, respectively, for classical conditioning of sensory-motor acts. The $\mathscr{A}_f^+$ cells are persistently turned on by shock (on-cells). The $\mathscr{A}_f^-$ cells are transiently turned on by shock termination (off-cells). The rebound from $\mathscr{A}_f^+$ cell activation to $\mathscr{A}_f^-$ cell activation replaces drive reduction in the case of shock. Classical conditioning from sensory cells $\mathscr{S}$ to the pattern of activity playing on arousal cells $\mathscr{A}_f = (\mathscr{A}_f^+, \mathscr{A}_f^-)$ can occur. Sufficiently positive net feedback from $\mathscr{A}_f$ to $\mathscr{S}$ can release sampling, and subsequent learning, by prescribed cells in $\mathscr{S}$ of motor output controls. Once sampled, these controls can be reactivated by $\mathscr{S}$ on recall trials. This concept avoids some difficulties of two-factor theories of punishment and avoidance.

Recent psychophysiological data and concepts are analyzed in terms of network analogs, and some predictions are made. The rebound from $\mathscr{A}_f^+$ cell activation to $\mathscr{A}_f^-$ cell activation at shock termination is interpreted to be a consequence of different rates of transmitter accumulation – depletion in the parallel neural channels associated with $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$. This interpretation culminates in an analogy with adrenergic and cholinergic interactions at lateral and ventromedial hypothalamic sites, dependent on phasic sensory input and tonic reticular formation input. Mechanisms are suggested for such phenomena as: the lesser rewarding effect of reducing J units of shock to $J/2$ units than of reducing $J/2$ units to 0 units; a relationship between the rewarding effect of reducing J units of shock to $J/2$ units and the possibility of releasing a conditioned avoidance response in the presence of fearful cues; two kinds of depressed emotional affect, one due to overarousal, that can also be associated with massive associational confusions and poor paying attention, and one due to underarousal, that can also be associated with overreactive fear and relief responses; persistent nonspecific fear that biases interpretation of specific cues, and can "resist" new learning or "repress" old learning; different effects of gradual and abrupt shock on response suppression; response generalization from one shock level to another; reduction of pain in the presence of loud noise (analgesia); influences of drugs, such as carbachol, atropine, and scopolamine on conditioned emotional and avoidance responses, and on self-stimulation via implanted

hypothalamic electrodes; sensory-drive heterarchy that allows changes in situational cues to release responses compatible with any of several nonprepotent drives; feedback inhibition of adrenergic transmitter production; potentiation of adrenergic production by presynaptic spiking, and by postsynaptic spiking via a feedback loop that controls higher-order instrumental conditioning; learning at cholinergic synapses.

1. INTRODUCTION

Part I of this article [29] derived qualitative neural mechanisms of punishment and avoidance from psychological postulates, and analyzed recent psychological data using these mechanisms. This part of the article derives quantitative versions of these mechanisms. Psychophysiological data will be discussed using the quantitative results, and some predictions will be made. These include:

1. a relationship between the reinforcing effect of reducing J units of shock to $J/2$ units of shock, and the possibility of eliciting a conditioned avoidance response in the presence of fearful cues;

2. a relationship between higher-order instrumental conditioning, and postsynaptic effects on presynaptic norepinephrine production via an anatomical feedback loop, which is possibly a formal analog of medial forebrain bundle pathways;

3. a formula to decide when reducing J_1 units of shock to K_1 units of shock in the presence of I_1 units of arousal is more reinforcing than reducing J_2 units of shock to K_2 units in the presence of I_2 units of arousal, where one arousal mechanism is a possible formal analog of reticular formation inputs;

4. a mechanism for two kinds of depressed emctional affect. The first kind, due to overarousal, is stable with respect to psychological inputs; the network is "indifferent" to emotionally charged cues. This effect can be associated with poor paying attention, associational confusions, and punning behavior. The second kind of depressed affect, due to underarousal, is an unstable form of depression in the sense that, after the system's elevated thresholds are exceeded by external cues, either aversive or rewarding cues can cause overreactive fear or relief responses; network response is emotionally "irritable";

5. a relationship between administration of carbachol (an acetylcholine mimicker) at lateral and ventromedial hypothalamic sites, and the spatial distribution of fearful and rewarding cues in the experimental chamber, to tell if carbachol will enhance or depress conditioned avoidance learning.

2. BRIEF REVIEW

Part I derives neural networks whose anatomy is of the type depicted in Fig. 1, where (Fig. 1a) the ith conditioned stimulus (CS_i) among n

possible stimuli excites state U_{i1} of its sensory representation. Sensory representations will be denoted generically by $\mathscr{S}$. In response to the CS_i input, U_{i1} sends signals to stage U_{i2} as well as to arousal cell clusters $\mathscr{A} = (\mathscr{A}_h, \mathscr{A}_f^+, \mathscr{A}_f^-, \ldots)$. The arousal cells $\mathscr{A}_h$ subserve hunger, $\mathscr{A}_f^+$ subserve fear, and $\mathscr{A}_f^-$ subserve relief from fear. The cells $\mathscr{A}_h$ receive a drive input that is a monotone increasing function of hunger level, and $\mathscr{A}_f^+$ receives an input that is a monotone increasing function of shock level. Offset of shock elicits a transient excitatory response in the relief

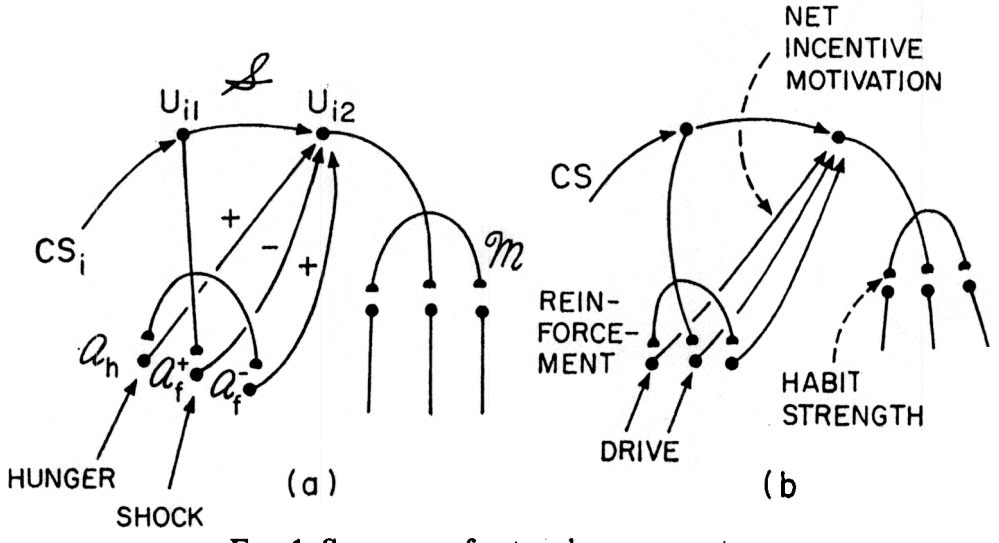

FIG. 1. Summary of network components.

center $\mathscr{A}_f^-$. Inputs from $\mathscr{A}_h$ and $\mathscr{A}_f^-$ to U_{i2} are excitatory, whereas the input from $\mathscr{A}_f^+$ to U_{i2} is inhibitory. U_{i2} can send signals to $\mathscr{M}$ only if it simultaneously receives a large signal from U_{i1} and a large excitatory signal from $\mathscr{A}$. In particular, a large excitatory $\mathscr{A}_h \rightarrow U_{i2}$ signal can be canceled by a large inhibitory $\mathscr{A}_f^+ \rightarrow U_{i2}$ signal, which thus prevents U_{i2} from firing. Moreover, if shock is terminated by an avoidance response (AR), $\mathscr{A}_f^-$ is excited and creates a large excitatory $\mathscr{A}_f^- \rightarrow \mathscr{S}$ input to all sensory representations. Feedback cues of the AR also excite particular sensory representations, which we denote by $\mathscr{S}$(AR). The U_{i2} stages of $\mathscr{S}$(AR) cells thus receive U_{i1} and $\mathscr{A}_f^-$ inputs. They can therefore fire. Cells that receive only the $\mathscr{A}_f^-$ input cannot fire.

Both $U_{i1} \rightarrow \mathscr{A}$ synaptic knobs and $U_{i2} \rightarrow \mathscr{M}$ synaptic knobs can learn ("sample") patterns of activity playing on $\mathscr{A}$ or $\mathscr{M}$, respectively, when these knobs are activated by U_{i1} or U_{i2} signals. The $U_{i1} \rightarrow \mathscr{A}$ synapses learn "motivational" patterns; the $U_{i2} \rightarrow \mathscr{M}$ synapses learn "motor" patterns playing on the motor control cells $\mathscr{M}$. Figure 1b describes the network components in more conventional psychological terms. Part I should be consulted for details.

The task of this article is to construct explicit networks for the rebound mechanism that transiently turns on the "relief" cells $\mathscr{A}_f^-$ when the shock input to the "fear" cells $\mathscr{A}_f^+$ is turned off.

Physiological laws for these networks have previously been derived [22, 26]. In their simplest form, these laws are

$$\dot{x}_i(t) = -\alpha_i x_i(t) + \sum_{k=1}^{n} [x_k(t - \tau_{ki}) - \Gamma_{ki}]^+ \beta_{ki} z_{ki}(t) - \sum_{k=1}^{n} [x_k(t - \sigma_{ki}) - \Omega_{ki}]^+ \gamma_{ki} + I_i(t) \tag{1}$$

$$\dot{z}_{jk}(t) = -\delta_{jk} z_{jk}(t) + \varepsilon_{jk}[x_j(t - \tau_{jk}) - \Gamma_{jk}]^+ x_k(t), \tag{2}$$

where $i, j, k = 1, 2, \ldots, n$ and $[\omega]^+ = \max(\omega, 0)$ for any real number ω. These laws describe interactions of stimulus traces (cell potentials) $x_i(t)$ at the cells (or cell clusters) v_i with memory traces (excitatory transmitter substances) $z_{jk}(t)$ at the synaptic knobs (or synaptic knob clusters) N_{jk} via the axons e_{jk}. The external event r_i excites the stimulus trace $x_i(t)$ via the input $I_i(t)$ (see Fig. 2). These equations are refined in [24]. The need for

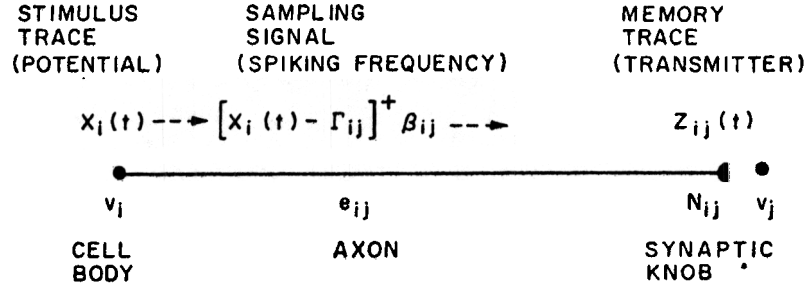

FIG. 2. Psychophysiological interpretation of network variables.

this refinement is seen by considering the terms $T_{ki}(t) = [x_k(t - \tau_{ki}) - \Gamma_{ki}]^+ \beta_{ki} z_{ki}(t)$ in (1). T_{ki} describes the rate of transmitter release from N_{ki}. The term $[x_k(t - \tau_{ki}) - \Gamma_{ki}]^+ \beta_{ki}$ is proportional to the spiking frequency reaching N_{ki} in the time interval $[t, t + dt]$, and $z_{ki}(t)$ is proportional to the amount of available transmitter in N_{ki} during $[t, t + dt]$. Given this interpretation of T_{ki}, why does not (2), which describes transmitter accumulation rate in N_{jk}, include an extra term $-T_{jk}(t)$ to describe release of transmitter? Such a term would destroy the memory of past pairings of the events r_j and r_k. This dilemma is resolved by using two variables $z_{jk}(t)$ and $Z_{jk}(t)$ instead of the single variable $z_{jk}(t)$. The new $z_{jk}(t)$ describes the rate of transmitter production. $Z_{jk}(t)$ describes the amount of available transmitter. The fact that $z_{jk}(t)$ and $Z_{jk}(t)$ can be lumped together in (1) and (2) means that $Z_{jk}(t)$ seeks a level proportional to $z_{jk}(t)$ that can transiently

be reduced by transmitter release. Thus (1) is replaced by

$$\dot{x}_i(t) = -\alpha_i x_i(t) + \sum_{k=1}^{n} [x_k(t - \tau_{ki}) - \Gamma_{ki}]^+ \beta_{ki} Z_{ki}(t) + I_i(t) \tag{3}$$

and

$$\dot{Z}_{jk}(t) = \zeta_{jk}(\eta_{jk} z_{jk}(t) - Z_{jk}(t)) - \kappa_{jk}[x_j(t - \tau_{jk}) - \Gamma_{jk}]^+ Z_{jk}(t). \tag{4}$$

Note that $Z_{jk}(t)$ replaces $z_{jk}(t)$ in (3). Equation (4) describes the net effect of three processes on rate of transmitter production: (i) transmitter production rate is proportional to $z_{jk}(t)$, as in the term $\zeta_{jk}\eta_{jk}z_{jk}(t)$; (ii) production rate is decreased by feedback inhibition proportional to the amount of transmitter, via the term $-\zeta_{jk}Z_{jk}(t)$; and (iii) amount of transmitter is reduced at a rate $\kappa_{jk}[x_j(t - \tau_{jk}) - \Gamma_{jk}]^+ Z_{jk}(t)$ by transmitter release. When no release occurs, $Z_{jk}(t)$ rapidly seeks the level $\eta_{jk}z_{jk}(t)$, proportional to $z_{jk}(t)$, as required by (1) and (2). In cases where no learning occurs, $\gamma_{jk} = \delta_{jk} = 0$ in (2); hence Z_{jk} seeks a constant level $\zeta_{jk}z_{jk}(0)$ in the absence of transmitter release. Some finer transients in transmitter production and release are also considered in [24].

3. REBOUND MECHANISM: IS IT UNIQUE?

How can a rebound mechanism be constructed in the context of Eqs. 2–4? We will impose increasingly sharp criteria on the mechanism until definite minimal constructions are suggested. Within the context of Eqs. 2–4, these minimal constructions will be essentially unique consequences of our criteria. There remains the possibility, however, that these equations have omitted mechanisms that can also generate a rebound from on-cells to off-cells. For example, consider Fig. 3: in Fig. 3a, a tonic input is applied equally to v_1 and v_2. Let this input be constant for simplicity. Shock also creates an input at v_1. Suppose that onset of shock creates an excitatory overshoot in x_1 and that offset of shock creates an inhibitory undershoot of x_1, as in Fig. 3b. After signals from v_1 and v_2 compete subtractively on their way to v_3 and v_4 in Fig. 1a (that is, form a subtractive on–off field), the net input to x_3 (the fear channel) is persistently turned on by shock, and the net input to x_4 (the relief channel) is transiently turned on by the inhibitory undershoot of x_1. Such overshoots and undershoots can be due (say) to ionic fluxes that can exist in a description of cell membrane dynamics which is finer than Eqs. 2–4. A rebound mechanism of this type is not chosen primarily because (i) such ionic rebounds are typically faster than the rebound process needed here. A rebound is needed that lasts on the order of seconds rather than milliseconds to enable conditioning in $U_{i1} \to \mathcal{A}$ and $U_{i2} \to \mathcal{M}$ channels to occur.

(ii) Also, the size of x_1's inhibitory undershoot should increase as a function of shock intensity, corresponding to data on the greater rewarding effects of terminating more intense shocks. In particular, the range of

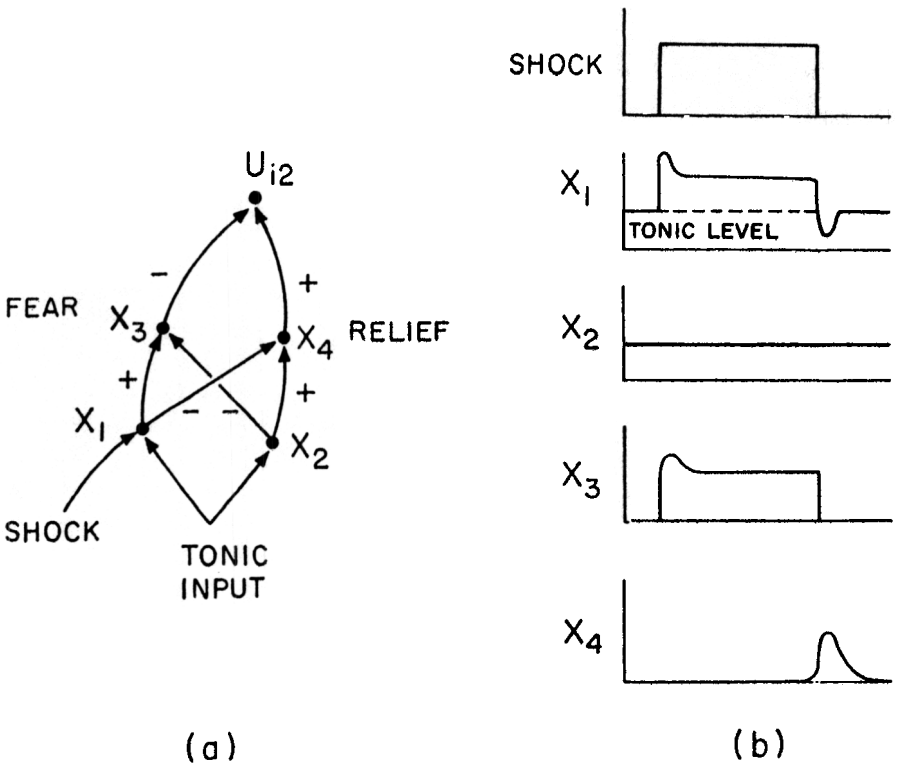

FIG. 3. Possible way to create relief after fearful shock terminates.

x_1's undershoot magnitudes as a function of shock intensity should be commensurate with the range of x_1's steady-state responses as a function of shock intensity (see Fig. 4); otherwise the parametric changes in undershoot will not generate significant changes in conditioned responding.

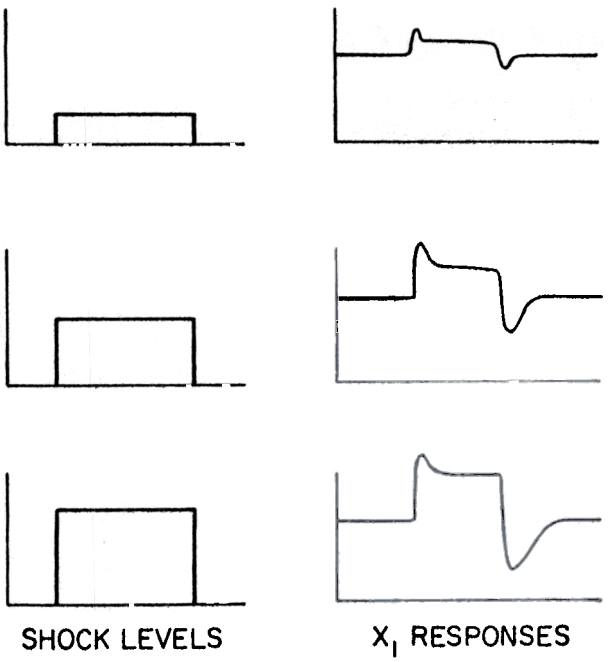

FIG. 4. Parametric dependence of potential response on shock intensity.

(iii) Moreover, the size of x_1's inhibitory undershoot should increase as a function of shock duration, corresponding to data on the greater rewarding effects of terminating shocks of longer duration. Thus the undershoot should have a memory, lasting in the range of seconds, of how long the shock was on.

(iv) Also, the size of x_1's excitatory response to shock should increase as a function of shock duration, corresponding to data on the greater fear created by longer shocks. Thus x_1 should gradually increase, after shock is turned on, to an asymptote that increases as a function of shock intensity. This property violates the simple overshoot–undershoot concept that is depicted in Fig. 3b.

If such difficulties as (i)–(iv) can be overcome by mechanisms other than those in Eqs. 2–4, then these mechanisms would also be candidates for a rebound mechanism. It presently seems difficult to overcome (i)–(iv) using ionic mechanisms. A mechanism using systems (2)–(4) is, however, readily derived, and does not require a finer description of cell dynamics. Henceforth, all arguments will be interpreted within the framework of this system.

4. REBOUND MECHANISM: HEURISTICS

The argument leading to the rebound mechanism falls into eight main stages. Some readers might prefer studying the mechanism in Section 5 before reading this section.

A. Existence of a Tonic Input

When shock terminates, $\mathscr{A}_f^-$ emits a transient output. Thus, by (3), the potentials x_f^- of $\mathscr{A}_f^-$ cells grow transiently to suprathreshold values. In (3), an input source is required to perturb x_f^- thus. What input source does the job? (The concept of "input source" includes possible energy sources within the cells themselves.)

In system (2)–(4), shutting off one input (such as the shock input to $\mathscr{A}_f^+$) does not provide energy for turning on another input (such as the one driving $\mathscr{A}_f^-$ rebound). Terminating shock input can, however, unmask the effects of an *internally* driven input to $\mathscr{A}_f^-$ whose influence is inhibited by shock. The internal source of $\mathscr{A}_f^-$ is therefore neither turned on nor turned off by shock offset. It is not turned off by shock onset, since then it would be off at shock offset, and could not drive $\mathscr{A}_f^-$ rebound. Finally, if it is turned on by shock onset, or is unaffected by shock onset, then it is always on. The internal input is therefore *tonic*.

B. Existence of Accumulation–Depletion

Output from $\mathscr{A}_f^-$ shuts off soon after it is turned on. How is this done? No externally driven input is available to do this. The $\mathscr{A}_f^-$ output is

depleted by its own activity. In other words, while shock is on, an accumulation process occurs at $\mathscr{A}_f^-$. When shock is off, output from $\mathscr{A}_f^-$ is a monotone increasing function of the amount accumulated at each given time. This amount is gradually depleted when shock is off, until the $\mathscr{A}_f^-$ output vanishes.

C. Consensus between Fear and Relief

We suppose that at most one of the outputs from $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ is nonzero at any time. In other words, either fear or relief, but not both, can be "perceived" by the network at a given time. Thus the final state of processing in $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$, before signals are sent to $\mathscr{S}$, is the resultant of a competition between the $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ channels due to some form of mutual inhibition, much as in Fig. 3a.

D. Existence of a Parallel Accumulation Process in the Fear Channel

When shock is off for a long time, outputs from both $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ to $\mathscr{S}$ are zero. Thus the accumulation process at $\mathscr{A}_f^-$, driven by its tonic input, is balanced by a process going on at $\mathscr{A}_f^+$. The simplest idea is that a parallel process of accumulation–depletion, driven by its own tonic input that equals the $\mathscr{A}_f^-$ input, takes place in the $\mathscr{A}_f^+$ channel. When shock is on, the shock input summates with the tonic input in the $\mathscr{A}_f^+$ channel.

This idea is strengthened by the next few arguments, which elucidate the basic question: what accumulates? Is it potential or is it transmitter? Several facts favor the latter alternative.

E. The Rebound is Slow

It lasts at least seconds rather than milliseconds. It is a slow process compared to known fluctuation rates of cell potentials in response to input changes. After shock terminates, $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ receive no externally driven inputs. Their potentials presumably rapidly equalize. Output from $\mathscr{A}_f^-$ nonetheless continues. Thus there exists a process slower than potential change that can bias output from $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ in favor of $\mathscr{A}_f^-$ after shock terminates.

F. Both Fear and Relief Are Increasing Functions of Shock Duration and Intensity

Data on the effect of CS and UCS intensity [2, 8, 17, 35–37, 45, 52] and duration [2, 6, 9, 14, 38, 58] on the CER and CAR have been reported. Thus both channels contain slowly varying processes which parametrically depend on shock intensity and duration, and which counterbalance each other when shock is off for long intervals.

G. The Relative Balance of Accumulation Is Changed by Shock

What causes the $\mathscr{A}_f^-$ rebound to shut itself off? Is complete depletion of the accumulated product at $\mathscr{A}_f^-$ responsible for this? Suppose "yes."

Then the tonic input alone can deplete $\mathscr{A}_f^-$. By symmetry, during shock, the shock input plus the tonic input to $\mathscr{A}_f^+$ could surely deplete $\mathscr{A}_f^+$. This does not occur, since then fear could not be maintained by a prolonged shock. A weaker conclusion is necessary: shock shifts the relative balance of accumulation in the two channels, by depleting the $\mathscr{A}_f^+$ channel more than the $\mathscr{A}_f^-$ channel.

H. Signal Size Is a Joint Function of Input Size and Amount Accumulated

This argument is crucial. During $\mathscr{A}_f^-$ rebound, both $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ receive equal tonic inputs which ultimately balance the amounts accumulated at $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$, and thereby nullify $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ signals to $\mathscr{S}$. Before this happens, $\mathscr{A}_f^-$ output exceeds $\mathscr{A}_f^+$ output because $\mathscr{A}_f^-$ accumulation exceeds $\mathscr{A}_f^+$ accumulation. In other words, given a fixed input size (the equal tonic inputs to $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$), output is an increasing function of accumulation level (in the two channels, $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$).

When shock is on, increasing shock intensity increases $\mathscr{A}_f^+$ output, since it causes an increase in fear. Increasing shock intensity also *decreases* the amount accumulated at $\mathscr{A}_f^+$; this is the basis of the rebound at $\mathscr{A}_f^-$ when shock is turned off. Thus, output is not a function of accumulation level alone, since then increasing shock intensity would decrease $\mathscr{A}_f^+$ output by decreasing the amount accumulated at $\mathscr{A}_f^+$. Output size is a joint function of input size and accumulation level.

By Eq. 3, output size is the product of spiking frequency and transmitter level. Spiking frequency is an increasing function of potential, which is an increasing function of input size. This leaves transmitter accumulation level as the abstract accumulation level discussed earlier. This argument commits us to our formalism. We could not proceed further unless (i) the amount of accumulated transmitter were a decreasing function of input size; and (ii) output size were nonetheless an increasing function of input size. Fortunately, both (i) and (ii) are true in system (2)–(4), and make our construction of the rebound mechanism possible in this context.

5. REBOUND MECHANISM: NONRECURRENT CASE

The minimal nonrecurrent (that is, feed-forward) embedding field compatible with Section 4 is defined as follows. Odd (even) subscripts denote cell sites associated with $\mathscr{A}_f^+(\mathscr{A}_f^-)$. Let

(a) v_1 and v_2 be the first stage of input processing;
(b) v_3 and v_4 be the second stage of processing;
(c) v_5 and v_6 be the third stage of processing;
(d) x_i be the potential of v_i, $\quad i = 1, 2, \ldots, 6$;
(e) z_1 and z_2 be the transmitters in N_{13} and N_{24}, respectively;
(f) $I(t)$ be the tonic internal input to v_1 and v_2;

(g) $J(t)$ be the phasic aversive input to v_1;
(h) O_i be the output of v_i, $i = 5, 6$.

The following process occurs in the network of Fig. 5.

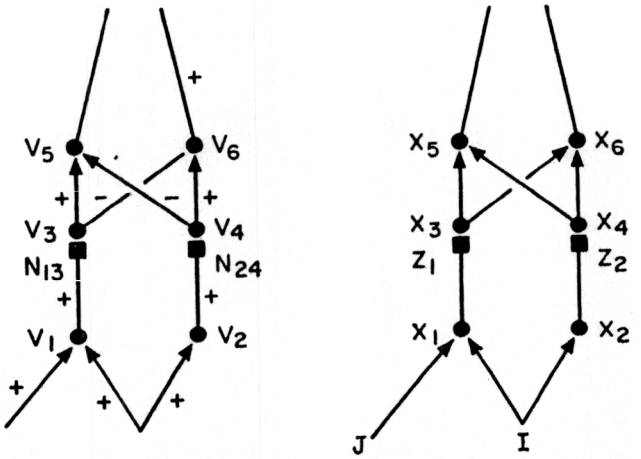

FIG. 5. Minimal nonrecurrent rebound mechanism using transmitter accumulation-depletion.

$$\dot{x}_1 = -\alpha x_1 + I + J, \tag{5}$$

$$\dot{x}_2 = -\alpha x_2 + I, \tag{6}$$

$$\dot{z}_1 = \beta(\gamma - z_1) - \delta[x_1(t-\tau) - \Gamma]^+ z_1, \tag{7}$$

$$\dot{z}_2 = \beta(\gamma - z_2) - \delta[x_2(t-\tau) - \Gamma]^+ z_2, \tag{8}$$

$$\dot{x}_3 = -\varepsilon x_3 + \zeta[x_1(t-\tau) - \Gamma]^+ z_1, \tag{9}$$

$$\dot{x}_4 = -\varepsilon x_4 + \zeta[x_1(t-\tau) - \Gamma]^+ z_2, \tag{10}$$

$$\dot{x}_5 = -\eta x_5 + \kappa[x_3(t-\sigma) - x_4(t-\sigma)], \tag{11}$$

$$\dot{x}_6 = -\eta x_6 + \kappa[x_4(t-\sigma) - x_3(t-\sigma)], \tag{12}$$

$$O_5 = \lambda[x_5 - \Omega]^+, \tag{13}$$

and

$$O_6 = \lambda[x_6 - \Omega]^+. \tag{14}$$

In (5), x_1 responds linearly to the sum $I + J$ of tonic and phasic inputs. In (6), x_2 responds linearly to the tonic input I alone. This asymmetry in inputs drives the outputs O_5 and O_6. The transmitter rules in (7) and (8) are of the form defined by (4) if no learning is possible. As in (4), the rate of transmitter change is controlled by three processes. For example, in (7), transmitter is produced at rate $\beta\gamma$. Feedback inhibition of transmitter production occurs at rate $-\beta z_1$. It is due to the action of transmitter, or a transmitter-activated substance, at an intermediate stage of transmitter production. Transmitter release from N_{13} occurs at rate $\delta[x_1(t-\tau) - \Gamma]^+ z_1$, where $[x_1(t-\tau) - \Gamma]^+$ is proportional to the spiking frequency from v_1 that reaches N_{13} at time t. When no spikes reach N_{13} in a given

interval, z_1 exponentially approaches a saturation level of γ at rate β. When spikes do reach N_{13}, z_1 is depleted at a rate that increases linearly with the spiking frequency. The same process determines accumulation and depletion of z_2 in N_{24}, except v_2 sends spikes to N_{24}. Thus asymmetries in the amounts of accumulated z_1 and z_2 can be traced, via the spiking frequency terms $[x_1(t-\tau)-\Gamma]^+$ and $[x_2(t-\tau)-\Gamma]^+$, to asymmetries in the inputs $I+J$ and I, respectively.

In (9), x_3 responds linearly to the signal $\zeta[x_1(t-\tau)-\Gamma]^+z_1$ from N_{13}, and in (10) x_4 responds linearly to the signal $\zeta[x_2(t-\tau)-\Gamma]^+z_2$ from N_{24}. The outputs from v_3 and v_4 are processed by a subtractive on–off field yielding inputs to v_5 and v_6. In (11), x_5 responds linearly to the difference $\kappa[x_3(t-\sigma)-x_4(t-\sigma)]$ of the signal from v_3 minus the signal from v_4. The output thresholds of v_3 and v_4 have been set equal to zero for simplicity. In (12), x_6 responds linearly to the difference $\kappa[x_4(t-\sigma)-x_3(t-\sigma)]$ of the signal from v_4 minus the signal from v_3. In particular, if the input to v_5 is positive, then the input to v_6 is negative, and conversely. This guarantees Property C of Section 4. The outputs (spiking frequencies) O_5 and O_6 from v_5 and v_6 in (13) and (14) are linear functions of the potentials x_5 and x_6 above the common spiking threshold Ω.

Equations 5–14 can be explicitly integrated step by step. Three phases of network activity are considered: (i) *Before shock*: Let $J(t)=0$ for $t \leqslant 0$. (ii) *During shock*: *Fix* $J(t)$ at a positive value J for $0 \leqslant t < T$. (iii) *After shock*: Let $J(t)=0$ for $t \geqslant T$. The following constraints are imposed. (A) Inputs I and J vary slowly relative to the fluctuation rate of potentials, except when J is switched on and off. (B) I is chosen sufficiently large to fire spikes along e_{13} and e_{24} even if $J=0$; positive J biases the pattern of firing. (C) Potentials adjust to input changes faster than transmitters accumulate. (D) T is sufficiently large to let transmitters adjust to the influence of positive J, and is large compared to the time lags τ and σ.

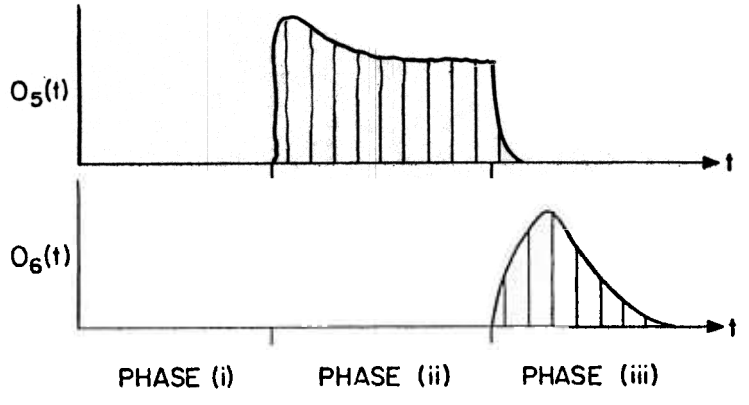

FIG. 6. Idealized persistent negative- and transient positive-incentive motivational responses to shock.

The outputs O_5 and O_6 will be studied during each phase. We will prove that in phase (i), $O_5 = 0 = O_6$. In phase (ii), O_5 becomes positive and approaches a positive asymptote, whereas $O_6 = 0$. O_5 is the basis for response suppression during phase (ii). In phase (iii), O_5 is rapidly driven to zero, but O_6 becomes positive and after achieving a positive asymptote decays to zero. O_6 during phase (iii) describes the rebound on which the CAR is based (see Fig. 6).

Consider phase (i). The inputs are symmetrically distributed; hence, the outputs are zero. To see this, note by (5) and (6) that $x_1 \cong \alpha^{-1}I \cong x_2$. Hence by (7) and (8), $z_1 \cong z_2$. Then by (9) and (10), $x_3 \cong x_4$, which in (11) and (12) yields $x_5 \cong 0 \cong x_6$. Finally, (13) and (14) imply $O_5 = 0 = O_6$.

Consider phase (ii). We compute the asymptote $x_5(T)$ of the $\mathscr{A}_f^+$ potential. $x_5(T)$ provides a measures of maximal fear and of the size of $\mathscr{A}_f^+ \to \mathscr{S}$ feedback. By Properties A and D, shortly after $t = 0$,

$$x_1 \cong x_1(T) \cong \alpha^{-1}(I + J) \tag{15}$$

$$x_2 \cong x_2(T) \cong \alpha^{-1}I. \tag{16}$$

Substitute (15) and (16) into (7) and (8). By Property B, $[\alpha^{-1}I - \Gamma]^+ > 0$. Thus, by (7) and Property A, z_1 is (approximately) a monotone decreasing function. By (8) and Property A, z_2 remains (approximately) constant. By Property D, z_1 and z_2 reach their new asymptotes before $t = T$. These asymptotes are

$$z_1(T) \cong \frac{\alpha\beta\gamma}{\alpha\beta + \delta(I + J - \alpha\Gamma)} \tag{17}$$

and

$$z_2(T) \cong \frac{\alpha\beta\gamma}{\alpha\beta + \delta(I - \alpha\Gamma)}. \tag{18}$$

Note that $z_1(T) < z_2(T)$: the aversive input depletes z_1 more than z_2. Nonetheless the signals $\zeta[x_1(t - \tau) - \Gamma]^+ z_1$ and $\zeta[x_2(t - \tau) - \Gamma]^+ z_2$ to v_3 and v_4 asymptotically satisfy the reverse inequality

$$\zeta[x_1(T - \tau) - \Gamma]^+ z_1(T) > \zeta[x_2(T - \tau) - \Gamma]^+ z_2(T).$$

This follows from (15)–(18) and Property D. By Property D, $x_i(T - \tau) \cong x_i(T)$. Thus by (15)–(18),

$$\zeta[x_1(T) - \Gamma]^+ z_1(T) \cong \frac{E(I + J - F)}{G + (I + J - F)} \tag{19}$$

and

$$\zeta[x_1(T) - \Gamma]^+ z_2(T) \cong \frac{E(I - F)}{G + (I - F)}, \tag{20}$$

with $E = \beta\gamma\delta^{-1}\zeta$, $F = \alpha\Gamma$, and $G = \alpha\beta\delta^{-1}$. The function

$$f(w) = \frac{E(w - F)}{G + (w - F)}$$

is monotone increasing in $w \geq F$ if E, F, and G are positive, so that (19) exceeds (20) if $J > 0$. Thus, the multiplicative coupling between spiking frequency and transmitter yields an output signal that is monotone increasing as a function of spiking frequency, even though the transmitter term is monotone decreasing as a function of spiking frequency. This is the crucial fact.

The asymptotic outputs O_5 and O_6 are found by substituting (19) and (20) into (9) and (10) and then $x_3(T)$ and $x_4(T)$ into (11) and (12). We find by using Property D that

$$x_5(T) \cong \frac{UJ}{(V + I)(V + I + J)}$$

and $x_6(T) \cong -x_5(T)$, with $U = \alpha\beta^2\gamma\delta^{-2}\zeta\varepsilon^{-1}\kappa\eta^{-1}$ and $V = \alpha(\beta\delta^{-1} - \Gamma)$. The following facts are obvious from (21).

PROPOSITION 1

For $I \geq \alpha\Gamma$, $x_5(T)$ is a monotone increasing function of J and a monotone decreasing function of I. Equations 13 and 21 show that $O_5(T)$ is a monotone increasing function of the aversive input J and that $O_6(T) = 0$. $O_5(T)$ is a monotone decreasing function of I. If the network $\mathcal{O}$ can control the size of I, then $\mathcal{O}$ can also reduce the CER created by an aversive input J of fixed size. The dependence of $O_5(T)$ on I will also be important when the maximum of O_5 in phase (ii) is compared with the maximum of O_6 in phase (iii). The ratio $O_5^{-1}(T) \max\{O_6(t): t \geq T\}$ controls the relative strength of motivational support for CAR and CER channels. The size of I influences this ratio.

To see this, consider phase (iii) using the approximation in which all potentials instantaneously adjust to the removal of J, whereas z_1 and z_2 gradually move from their asymptotes $z_1(T)$ and $z_2(T)$ to their new asymptotes. Let the notation $f(T+)$ denote the (approximate) value of the function f at a time, shortly after $t = T$, when the x_i's have readjusted but the z_i's have not. We study $x_6(T+)$, which is the maximum of the rebound at x_6 during phase (iii), given instantaneous adjustment of potentials.

First note that

$$x_1(T+) \cong \alpha^{-1}I \cong x_2(T+)$$

whereas

$$z_1(T+) \cong z_1(T) < z_2(T) \cong z_2(T+).$$

At $t \cong T+$, the signals $\zeta[x_1(t-\tau)\Gamma]^+ z_1$ and $\zeta[x_2(t-\tau)-\Gamma]^+ z_2$ from N_{13} and N_{24} satisfy the inequality

$$\zeta[x_1(T+)-\Gamma]z_1(T+) < \zeta[x_2(T+)-\Gamma]z_2(T+), \tag{24}$$

since by (17), (18), and (22),

$$\zeta[x_1(T+)-\Gamma]z_1(T+) \cong \frac{E(I-F)}{G+(I+J-\Gamma)} \tag{25}$$

$$\zeta[x_2(T+)-\Gamma]z_2(T+) \cong \frac{E(I-F)}{G+(I-F)}. \tag{26}$$

Inequality (24) is the basis of the rebound at v_6. Substituting (25) and (26) into (9) and (10) and then (11) and (12) yields

$$x_6(T+) \cong \frac{WJ(I-F)}{(V+I)(V+I+J)}, \tag{27}$$

and $x_5(T+) \cong -x_6(T+)$, with $W = \beta\gamma\delta^{-1}\zeta\varepsilon^{-1}\kappa\eta^{-1}$. A routine computation proves the following facts.

PROPOSITION 2

$x_6(T+)$ is a monotone increasing function of J. $x_6(T+)$ is monotone increasing in I for $F \leqslant I \leqslant F + (G^2 + JG)^{\frac{1}{2}}$ and monotone decreasing in I for $I > F + (G^2 + GJ)^{\frac{1}{2}}$. The maximum

$$\frac{WJ(G^2+JG)^{\frac{1}{2}}}{[G+(G^2+JG)^{\frac{1}{2}}][J+G+(G^2+JG)^{\frac{1}{2}}]}$$

of $x_6(T+)$ as a function of I, for fixed J, is a monotone increasing function of J. Thus there exists a value of I that maximizes the rebound driving the CAR for fixed J. Both this maximum and $x_6(T+)$ for fixed I are monotone increasing in J.

The following proposition studies the CAR/CER ratio, which estimates the relative size of O_6 and O_5 during phase (iii) and phase (ii), respectively. The importance of this ratio is shown by the following example. Let the rebound at O_6 be small relative to the size of O_5 during shock; that is, CAR/CER $\ll 1$. Let avoidance cues and cues conditioned to response suppression be simultaneously present during avoidance trials. Then the avoidance cues cannot provide enough positive feedback to $\mathscr{S} \to \mathscr{M}$ channels to overcome the CER and release the CAR. A large CAR/CER ratio means that avoidance cues can, in principle, overcome fear and induce a CAR. We study the ratio $x_5^{-1}(T)x_6(T+)$ instead of the closely related ratio $O_5^{-1}(T)\max\{O_6(t): t \geqslant T\}$.

PROPOSITION 3

$$x_5^{-1}(T)x_6(T+) = G^{-1}(I - F). \tag{28a}$$

Thus

$$x_5^{-1}(T)x_6(T+) = \left(1 + \frac{J}{G}\right)^{\frac{1}{2}} \quad \text{if} \quad I = F + (G^2 + JG)^{\frac{1}{2}}. \tag{28b}$$

Note in (28a) that the CAR/CER ratio is independent of J. In other words, the tonic arousal level determines the *relative* preference for relief over fear in the rebound mechanism. By (28a), increasing I biases the CAR/CER ratio in favor of avoidance. By (21) and (27), however, increasing I to large values drives O_5 and O_6 toward zero, and therefore eliminates both negative and positive $\mathscr{A}_f^{\pm} \to \mathscr{S}$ feedback; all "emotional" support of responding is removed. Nonetheless, letting $I = F + (G^2 + JG)^{\frac{1}{2}}$ maximizes the rebound $O_6(T+)$, by Proposition 2, and yields a CAR/CER ratio that favors avoidance, by (28b), by an amount that increases like the square root of J. Thus there exist values of I which create both a favorable CAR/CER ratio and sufficient arousal to drive the CAR.

6. ANALGESIA, DEPRESSION, NOVELTY, AND INVERTED U IN LEARNING

Equation (27) estimates the rebound created by turning off J units of shock. Denote this rebound by $\{J \to 0\}$. Equation (27) shows that the rebound caused by shutting off J units is greater than that caused by shutting off $J/2$ units; that is, $\{J \to 0\} > \{J/2 \to 0\}$. We now show that, moreover, $\{J/2 \to 0\} > \{J \to J/2\}$. This inequality can be interpreted as follows. Suppose that two aversive input sources summate at v_1. Let each source create $J/2$ units of input. Then the chain of inequalities

$$\{J \to 0\} > \left\{\frac{J}{2} \to 0\right\} > \left\{J \to \frac{J}{2}\right\}$$

means that it is most reinforcing to shut off both aversive sources; second most reinforcing to shut off one source in the absence of the other; and least reinforcing to shut off one source and leave the other on. This result illustrates the importance of parallel input channels on perceived fear and relief. Various pain analgesic effects [12, 20, 48] also suggest such influences. For example, loud white noise attenuates perceived pain due to drilling by a dentist. The attenuating influence of nonspecific inputs I to both $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ on perceived fear, as in (21), should be noted in this connection.

PROPOSITION 4

$\{J/2 \rightarrow 0\} > \{J \rightarrow J/2\}$. A rebound exists at v_6 in the case $\{J \rightarrow J/2\}$ if and only if $G^{-1}(I - F) > 1$; that is, if and only if the CAR/CER ratio favors avoidance. Otherwise, the transition $\{J \rightarrow J/2\}$ merely results in a reduction in fear.

PROOF

For convenience of exposition, let $x_5(T, J \rightarrow K)$ and $x_6(T+, J \rightarrow K)$ denote the fear and relief maxima created by switching J units of shock that are on throughout the interval $[0, T)$ to K units of shock at time T. A routine computation shows that for the transition $\{J \rightarrow J/2\}$,

$$x_6\left(T+, J \rightarrow \frac{J}{2}\right) = \frac{W(J/2)(I - F - G)}{(V + I)(V + I + J)}, \tag{30}$$

which is positive if and only if $G^{-1}(I - F) > 1$.

By (28a) and (30), $x_6(T+, J \rightarrow J/2) > 0$ if and only if, for any $K > 0$, $x_6(T+, K \rightarrow 0) > x_5(T, K \rightarrow 0)$. Indeed, by (28a) and (30), for all J, $K > 0$ and $I \geqslant \alpha\Gamma$,

$$x_6\left(T+, J \rightarrow \frac{J}{2}\right) = \frac{WG(J/2)[x_5^{-1}(T, K \rightarrow 0)x_6(T+, K \rightarrow 0) - 1]}{(V + I)(V + I + J)} \tag{31}$$

In any case,

$$x_6\left(T+, J \rightarrow \frac{J}{2}\right) < x_6\left(T+, \frac{J}{2} \rightarrow 0\right),$$

since by (27),

$$x_6\left(T+, \frac{J}{2} \rightarrow 0\right) = \frac{W(J/2)(I - F)}{(V + I)[V + I + (J/2)]},$$

which is larger than (30). ■

Equation 31 relates the amount of reward due to cutting J units of shock in half to the relative strength of fear and relief channels. Such a relationship is, in principle, testable and predicts that an individual who finds cutting shock in half unusually unrewarding should also have difficulty performing avoidance tasks in the presence of fearful cues.

Propositions 2 and 4 are special cases of the following general principle.

THEOREM 1 (*Reinforcement Ordering*)

Let two copies $\mathcal{N}_1$ and $\mathcal{N}_2$ of the network in Fig. 5 be given. Let the tonic input in $\mathcal{N}_i$ have size $I_i \geqslant F$. Let an aversive input of size J_i be switched on in $\mathcal{N}_i$ for a duration of T time units. Then let the aversive input level be

switched to K_i, $i = 1, 2$. Switching from J_1 to K_1 is more reinforcing than switching from J_2 to K_2 if and only if

$$R(I_1, J_1, K_1) > R(I_2, J_2, K_2), \tag{33}$$

where

$$R(I, J, K) = \frac{W(J - K)(I - F) - WKG}{(V + I)(V + I + J)}$$

Theorem 1 breaks up the total input to v_1 into two parts: an input I to v_1 and v_2 equally, which fluctuates slowly if at all; and an input J to v_1 alone, which can be quickly switched between various levels. In later sections, another source of inputs will be applied to v_1 and v_2; namely, the $\mathscr{S} \to \mathscr{A}_f^{\pm}$ input channels. At various times, some $\mathscr{S} \to \mathscr{A}_f^{\pm}$ channels will contribute equal inputs to $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ ("irrelevant cues"), some $\mathscr{S}$ channels will contribute a larger input to $\mathscr{A}_f^+$ than to $\mathscr{A}_f^-$ ("secondary aversive cues"), and some $\mathscr{S}$ channels will contribute larger inputs to $\mathscr{A}_f^-$ than to $\mathscr{A}_f^+$ ("secondary rewarding cues"). Because inputs summate at $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$, the crucial quantity is the relative size of the *total* inputs to $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$. If, for example, the total input to $\mathscr{A}_f^+$ exceeds that to $\mathscr{A}_f^-$, then the net effect is fear, even if several $\mathscr{S}$ channels project primarily to $\mathscr{A}_f^-$. If the total rapidly drops at $\mathscr{A}_f^+$, the net effect is a rebound at $\mathscr{A}_f^-$. The size of the rebound depends, as (34) shows, on both the absolute drop size and the size of the total "irrelevant" input at $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$.

Both unduly small and unduly large I levels create "abnormal" learning by "emotionally depressing" network response to phasic inputs, and thereby removing the incentive motivation that controls $\mathscr{S} \to \mathscr{M}$ sampling. Thus an "inverted U" in learning exists [34]. To study the inverted U, set I at successively higher values, starting with zero. If $I < F$, no rebound is possible, since by (6), (8), and (10), no signals pass from v_2 to v_4. Also, the fear threshold is raised, in the sense that a larger value of J is needed to create signals from v_1 to v_3 than is needed when $I \geqslant F$: neither fear nor relief occur in response to J inputs of "normal" size. If $I \gg 0$, the network is again "emotionally depressed," but for a different reason. A fixed J level, or secondary aversive input, or secondary rewarding input, creates a relatively small asymmetry in the total inputs to v_1 and v_2. Thus both O_5 and O_6 are very small, as (21) and (27) show. Any unduly large source of equal inputs to v_1 and v_2 will have this effect.

The two types of depression can be distinguished by increasing the J level step by step. In the case where $I < F$, the fear threshold is high, but once J is sufficiently large to fire $v_1 \to v_3$ signals, then additional increments in J create *larger* than normal increments in fear ("hyperexcitable" with high threshold). In fact, for any $I \geqslant 0$, $\partial x_5(T)/\partial J = U(V + I + J)^{-2}$,

which is quadratically decreasing in I. In a similar fashion, small rewarding inputs create larger than normal increments in relief. In the case where $I \gg 0$, the fear threshold is low, but increments in J create abnormally small increments in fear ("hypoexcitable" with low threshold). In this sense, the overaroused form of reduced affect is insensitive (is "indifferent") to emotionally charged events, whereas the underaroused form of reduced affect can overreact (is "irritable") to emotionally charged events that exceed the abnormally large threshold of this system.

In a similar fashion, either strongly conditioned $\mathscr{S} \to \mathscr{A}_f^+$ $(\mathscr{A}_f^-)$ channels that are persistently turned on, or an overaccumulation of transmitter at $\mathscr{A}_f^+$ $(\mathscr{A}_f^-)$ synaptic knobs, can produce persistent fear (relief). In the overaccumulation cases, these persistent fear or relief reactions are independent of the occurrence of particular cues, and therefore form a general emotional context which biases the interpretation of specific cues. In the case of strongly conditioned $\mathscr{S} \to \mathscr{A}_f^+$ channels, the negative feedback $\mathscr{A}_f^+ \to \mathscr{S}$ inhibits both learning of new $\mathscr{S} \to \mathscr{M}$ patterns and performance of old $\mathscr{S} \to \mathscr{M}$ patterns in response to the cues that elicit $\mathscr{S} \to \mathscr{A}_f^+$ signals. These effects can create a "resistance" to unlearning the fearful "meaning" of these cues, and a "repression" of the patterns encoded in their $\mathscr{S} \to \mathscr{M}$ synapses [27].

Another possible cause of the inverted U in learning has been studied. Grossberg and Pepe [30, 31] discuss the influence of $\mathscr{A} \to \mathscr{S}$ overarousal on serial learning by $\mathscr{S}$. Pathologically small $\mathscr{S} \to \mathscr{S}$ spiking thresholds or inadequate $\mathscr{S} \to \mathscr{S}$ lateral inhibition produce similar effects. They prove that overarousal yields associational confusions between erroneous choices that are closely related, in time or meaning, to the correct association. These "fuzzy" associational sets interfere with paying attention and subsequent learning. In this model, overarousal influences the relative strength of associations at the beginning and end of a serially learned list (primacy/recency ratio), the skewing of the bowed serial position curve, and the distribution of remote errors. Some of these effects seem not to have been studied experimentally, and if true would provide a conceptual bridge between the effects of overarousal on paying attention and other serial learning parameters.

The two inverted U's are due to different network mechanisms. The inverted U that yields depressed emotional affect will, in Section 9, be compared to the response of midbrain reward and punishment centers, such as hypothalamic sites, to variations of various arousal parameters. The inverted U that yields poor paying attention and associational confusions can be compared to responses of neocortical cells to variations in arousal level. The reticular formation supplies arousal to both regions. Our results therefore suggest that reticular formation overarousal can

yield a behavioral syndrome combining poor attention, massive associational confusions, and depressed emotion affect. Other type of arousal deficits can yield other symptoms. These will be compared with clinical findings in another place.

The size of I must be carefully controlled to avoid emotional under- or overarousal. What cells control the internal tonic part of I? Possibly v_1 and v_2 contain endogenous energy sources. If not, a likely interpretation of the formal input source is the ascending reticular activating system [59, page 434]. Kornetsky and Eliasson [42] and Phillips and Bradley [51] report that under- and overarousal of the reticular formation yields an inverted U. Can reticular overarousal also yield "hypoexcitable emotional depression"? The total uniform $\mathscr{S} \rightarrow \mathscr{A}_f^{\sharp}$ input also contributes to I and is therefore also carefully controlled. Section 10 discusses another reason and a mechanism for doing this.

Given such diverse sources of I input, various subtle interactions can occur. If the "novelty" or abruptness of shock creates a brief decrease in total I, then by (21), more fear will follow abrupt shock than shock which is gradually switched on [13, 47]. Grossberg [25, 28] discusses some anatomies that can discriminate complex input characteristics, including velocity detectors. An excitatory transient in J due to abrupt shock can also produce more fear, by (21). Grossberg [24] describes such a transient due to prior transmitter accumulation. A sufficiently slow overshoot of the type in Fig. 3b could also have this effect. New $\mathscr{S}$ channels can also be switched on or off by rate differences in shock onset. Suppose gradual shock causes less initial response suppression than abrupt shock. Then more motor activity R is possible and more sensory representations $\mathscr{S}$, denoted by $\mathscr{S}(R)$, which are activated by feedback cues of R, can be conditioned, either to shock ("spreading the fear around $\mathscr{S}$") or to unsuppressed consummatory activity (counterconditioning). The latter conditioning can produce resistance to the suppressive effect of a later intense shock. Suppose that shock also has an $\mathscr{S}$ representation; that is, the sensory channels activated by a given shock level J_1 also project to the first stage of sensory processing in certain $\mathscr{S}$ representations denoted by $\mathscr{S}(J_1)$, and these $\mathscr{S}(J_1)$ channels can sample $\mathscr{A}_f^{\sharp}$. Then $\mathscr{S}(J_1)$ sites can be conditioned to the prevailing motor activity, whether it is response suppression, given intense shock, or consummatory responding, given weak shock. If some $\mathscr{S}(J_1)$ channels are activated by a different shock level J_2 on recall trials, then the motor activity prevalent on learning trials given shock level J_1 can generalize to the J_2 level [13].

The foregoing discussion tacitly answers the question: What cell sites control "fear" and "relief" in Fig. 5? If fear and relief are mutually exclusive attributes, they are "perceived" after the operation of the

subtractive on–off field. In Fig. 5, this means that "fear" is controlled by v_5 and "relief" by v_6.

7. CONDITIONING FEAR AND RELIEF

No learning occurs in Fig. 5. $\mathscr{S}$ and $\mathscr{M}$ must be included in the figure. Where do $\mathscr{S}$ axons terminate? They terminate after the rebound mechanism, to permit CER and CAR learning, and before the on–off field, since fear and relief are mutually exclusive. Thus $\mathscr{S}$ projects to v_3 and v_4, as in Fig. 7, which depicts synaptic knobs at which learning occurs as semicircles, knobs at which (slow) accumulation takes place as squares, and all other knobs as arrowheads. Two formally distinct transmitter processes now converge at v_3 and v_4.

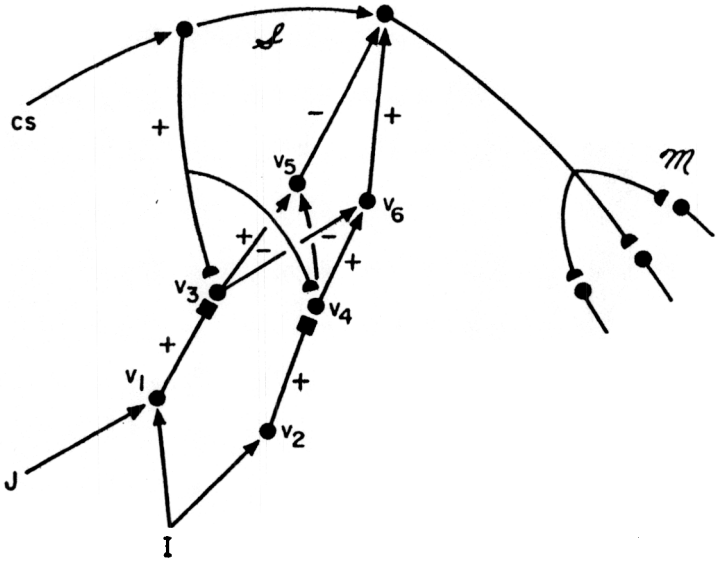

FIG. 7. Minimal synthesis of sampling and accumulation-depletion mechanisms.

Higher-order "classical" fear or relief conditioning is possible in Fig. 7. If $\mathscr{S}_1$ projects strongly to v_3 (v_4) and $\mathscr{S}_2$ projects equally to v_3 and v_4, then simultaneous activation of $\mathscr{S}_1$ and $\mathscr{S}_2$ will strengthen the $\mathscr{S}_2$ connection to v_3 (v_4), and weaken, or extinguish, the $\mathscr{S}_1$ connection to v_3 (v_4). Higher-order "instrumental," or rebound, conditioning does not occur in Fig. 7. Cessation of CS_1, when $\mathscr{S}_1$ projects strongly to v_3, does not drive a rebound at v_4, because accumulation occurs at the stage prior to the v_3 and v_4 cell sites. The next section considers the minimal extension of Fig. 7 in which higher-order instrumental conditioning is possible.

8. HIGHER-ORDER INSTRUMENTAL CONDITIONING

This section implements the following postulate.

Postulate XII. Higher-order instrumental conditioning is possible.

By Postulate XII, the network will be extended so that $\mathscr{S}$ can drive a conditioned rebound from $\mathscr{A}_f^+$ to $\mathscr{A}_f^-$. Thus $\mathscr{S}$ will send axons to a stage *prior* to the rebound, and these axons will have conditionable synaptic knobs. By Fig. 7, $\mathscr{S}$ sends axons to a stage *after* the rebound. Thus the anatomy of $\mathscr{A}_f$ is *recurrent*. The cell sites v_1 and v_3 are identified, as are v_2 and v_4. See Fig. 8 for some recurrent anatomies: In Fig. 8a, termination

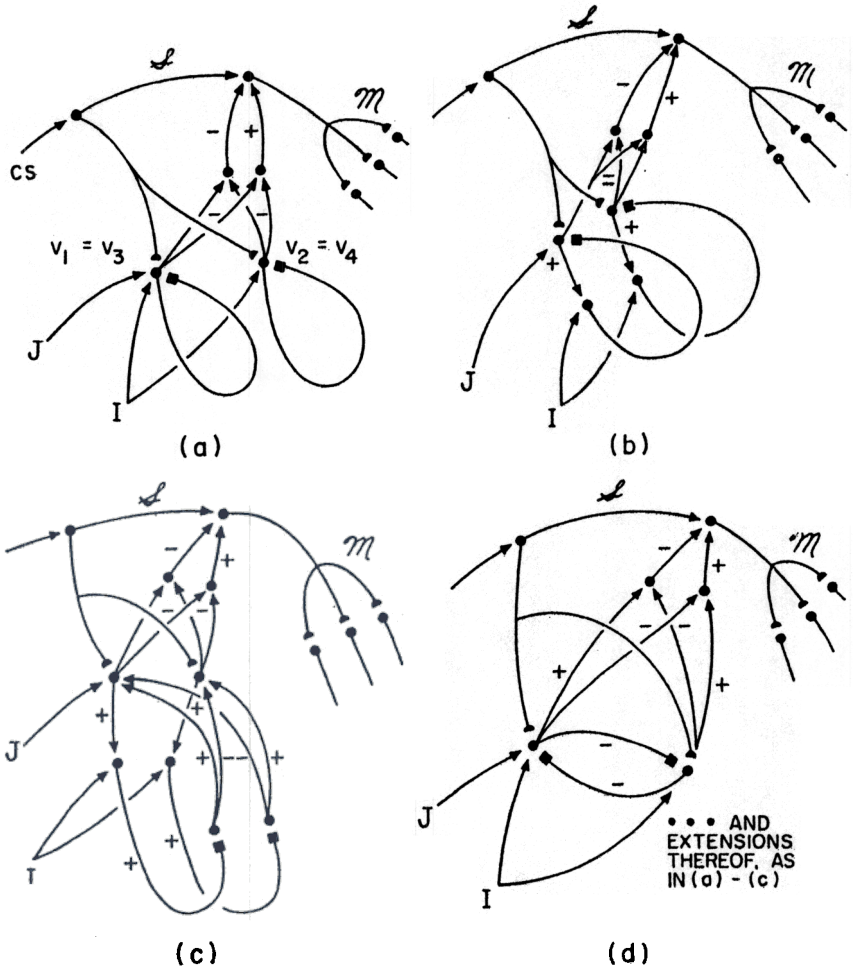

FIG. 8. Some recurrent accumulation-depletion schemes.

of J or of a conditioned aversive input from $\mathscr{S}$ to v_1 can drive a rebound at v_2. Thus higher-order instrumental conditioning is possible. Some important properties of this anatomy are listed below. Termination on a conditioned rewarding input from $\mathscr{S}$ to v_2 can also drive a rebound at v_1, yielding a brief increment in fear after abrupt termination of an interval of relief. This second-order fear conditioning effect is relatively small. Conditioned inputs from $\mathscr{S}$ to v_1 or v_2 are actively extinguished by the positive equal I inputs to v_1 and v_2 on recall trials. The recurrent "accumulation loops" $v_1 \rightarrow v_1$ and $v_2 \rightarrow v_2$ distort the relative intensity of the total $\mathscr{S}$,

J, and I inputs to v_1 and v_2 in the direction of "contour enhancement"; they produce a "contrast" effect (cf. [23]).

Figures 8b and 8c extend Fig. 7a in a way that diminishes the extinction on recall trials of conditioned $\mathscr{S} \rightarrow v_1$ and $\mathscr{S} \rightarrow v_2$ pathways due to equal I inputs at v_1 and v_2. In Fig. 8b, I only extinguishes $\mathscr{S}$ channels via the recurrent $v_1 \rightarrow v_1$ and $v_2 \rightarrow v_2$ loops, rather than directly and recurrently as in Fig. 8a. In particular, if a given $\mathscr{S}$ representation projects strongly to v_1 on recall trials, then the feedback along $v_1 \rightarrow v_1$ and $v_2 \rightarrow v_2$ will still favor v_1 in spite of I. $\mathscr{S}$ never samples the uniformly distributed I input, although the I input does indirectly reduce the relative strength of the $v_1 \rightarrow v_1$ feedback channel. In Fig. 8c, the effect of I on extinction is largely canceled out by the subtractive on–off field juxtaposed between the accumulation step and v_1 and v_2. These inhibitory interneurons cancel out the uniform part of the input to them, including the I input. $\mathscr{S}$ samples only the non-uniform part of the feedback. Extinction of $\mathscr{S} \rightarrow \mathscr{A}_f$ channels is due primarily to the influence of "irrelevant" $\mathscr{S}$ cues, namely, $\mathscr{S}$ inputs that are equal at $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$, or to counter conditioning by aversive cues. In this anatomy, therefore, a CAR is very stable if the CAR removes $\mathscr{O}$ from the cues driving the CAR itself, or at least from the cues of the aversive situation. Extinction of the CAR is possible if aversive cues are not terminated; cf. one-way versus two-way avoidance tasks.

Figure 8d describes a recurrent inhibition-with-accumulation between v_1 and v_2. The accumulating transmitter is here an inhibitory transmitter rather than an excitatory transmitter, as in Figs. 8d–c. Thus, if v_1 receives a larger input than v_2, then v_1 inhibits v_2 more than conversely, and transmitter is depleted more in N_{12} than in N_{21}. When the inputs to v_1 and v_2 equalize, then $v_2 \rightarrow v_1$ inhibition exceeds $v_1 \rightarrow v_2$ inhibition. Hence the output from v_2 exceeds that from v_1, and a rebound occurs.

A useful exercise for the reader is to perform "ablation" and "nerve section" experiments on these networks at various loci and to check what happens to the fear, relief, and rebound reactions.

9. CHOLINERGIC AND ADRENERGIC TRANSMITTERS

Our networks contain two formally distinct transmitter systems which converge at $\mathscr{A}_f^{\pm}$ sites to mediate the CER and CAR: learning occurs at semicircular synapses and slow accumulation–depletion occurs at square synapses. *In vivo*, two distinct transmitter systems influence CER and CAR learning and performance at hypothalamic sites, namely, the cholinergic and adrenergic systems. This striking coincidence is interpreted below. The interpretation is necessarily tentative, since Fig. 8 describes, at best, one fragment of the several systems needed during a routine learning task. This description is also minimal and presumably does not encompass all

the tasks addressed by the analogous system *in vivo*. Other systems also contain cholinergic and adrenergic components [56], and Fig. 8 provides little insight into their relative importance. Moreover, even if in the minimal description two or more synaptic knobs end at a given cellular site, which we interpret (say) as analogous to a lateral hypothalamic site, the corresponding system *in vivo* might separate these knobs by interneurons; cf. Fig. 8a–c. Hence our model might predict the existence of both cholinergic and adrenergic interactions at a single hypothalamic site, whereas *in vivo* these interactions might be occurring in parallel at different hypothalamic, or even nonhypothalamic, sites that are separated by interneurons. Nonetheless, the analogy suggests so many precise questions, possible insights concerning the relevant data, and directives for further studies that it will be given here. Some questions will be mixed with interpretive comments below.

Interpret the semicircular synapses as cholinergic synapses; they contain acetylcholine (ACh). Let the square synapses be adrenergic; they contain norepinephrine (NE). Let $\mathscr{A}_f^+$ be associated with ventromedial hypothalamus (VMH). Let $\mathscr{A}_f^-$ be associated with lateral hypothalamus (LH). Given this labeling, then both ACh and NE influence both VMH and LH sites.

Some encouraging data are readily noticed. Central monoamine neurons are tonically active [1], as are square synapses driven by the tonic internal I input. Is the *in vivo* tonic activity inhibited by blockade of reticular formation activity, or is it endogenous? Can over- and underarousal due to this tonic activity cause emotional depression?

We now note that all the terms in Eq. (7) are qualitatively supported by data on NE regulation. The rate of NE synthesis is controlled by the amount of NE; that is, feedback inhibition of endogenous monoamine biosynthesis exists [5]. This feedback inhibits the enzyme tyrosine hydroxylase, which controls the rate-limiting step, from tyrosine to DOPA, in NE production [60, 61]. Feedback inhibition of z_1 production exists in Eq. (7), in the term $-\beta z_1$. Increased adrenergic nervous activity yields enhanced synthesis of NE from tyrosine, which is prevented by adding NE to the bath, in the isolated hypogastric nerve–vas deferens preparation of the guinea pig [60]. The terms $-\beta z_1$ and $-\delta[x_1(t-\tau) - \Gamma]^+ z_1$ in Eq. (7) have a similar effect. The term $\delta[x_1(t-\tau) - \Gamma]^+ z_1$ depletes transmitter by releasing it at a rate that increases with presynaptic spiking frequency. Feedback inhibition of transmitter production is thereby reduced. Adding transmitter increases the term $-\beta z_1$ and thereby increases feedback inhibition of transmitter production. All terms in Eq. (7) are thus qualitatively supported by data. A model of some faster transients in transmitter production is considered in [24].

NE often, but not always, has inhibitory effects on postsynaptic sites [5]. In Figs. 8a–c, the square synapses are excitatory. Nonetheless, if NE is nonspecifically applied to the $\mathscr{A}_f^{\pm}$ region, and acts equally on $v_1 \rightarrow v_1$ and $v_2 \rightarrow v_2$ loops, then the *outputs* from $\mathscr{A}_f$ to $\mathscr{S}$ can be inhibited by the subsequent increase in total I input. In the recurrent anatomy of Fig. 8d square synapses are inhibitory. Bloom and Giarman [5] discuss recurrent inhibitory NE pathways in the olfactory bulb. Is Fig. 8d an analog of some such pathways in a different anatomical locus? Can a rebound be elicited at suitable olfactory sites?

The olfactory example suggest a basic question: Why are off-cells so prevalent in sensory processing regions [59, pages 253, 349]? One formal answer seems clear in these networks: Suppose $\mathscr{O}$ can learn that offset of a given stimulus is the cue for a given response. Then cells which are reliably turned on by stimulus offset, and which can sample $\mathscr{A}$ and $\mathscr{M}$, are needed. These off-cells would appear in $\mathscr{S}$, and would be driven by the on-cells that respond to the stimulus. The stimulus replaces the aversive input J in this case. The "dipole" of on-cell and off-cell can thus discriminate both externally important cues in $\mathscr{S}$ and internally important cues in $\mathscr{A}$; the different roles for such dipoles need depend only on the interpretation of what input source drives the on-cell and on the interpretation of what cells receive the on-cell and off-cell signals. A tonic internal input would be needed to drive the dipoles in $\mathscr{S}$, just as it was needed in $\mathscr{A}$. Do such dipoles exist in cerebral cortex? The profound influence of the ascending reticular activating system on cerebral cortex is well known [59, Chapter 14]. Are cerebral dipoles, supposing that they exist, driven by tonic reticular input or by endogeneous energy sources? Are cerebral dipoles adrenergic, or are they of the type described by Fig. 3?

Reserpine depletes NE, thereby reducing both sensitivity to shock and lever-press shock avoidance responses [55, 61]. Is a formal analog of this depletion a reduction in transmitter accumulation in the $v_1 \rightarrow v_1$ and $v_2 \rightarrow v_2$ synaptic knobs, and thus of the signals driving $\mathscr{A}_f^+$ and $\mathscr{A}_f^-$ output? Self-stimulation behavior is also dependent on NE activity [62]. Is this formally analogous to varying the amount of feedback driven by the stimulating electrode in the $v_1 \rightarrow v_1$ pathway by altering the amount of accumulation in $v_1 \rightarrow v_1$ synaptic knobs? Increased shock intensities produce more rapid recovery from reserpine [61]. Is this formally analogous to a greater synthesis of transmitter driven by higher presynaptic spiking frequencies? Medial forebrain bundle (MFB) lesions in the LH cause increased sensitivity to shock [32]. Is this formally analogous to removing $\mathscr{A}_f^-$ inhibition of $\mathscr{A}_f^+$ output, and thereby generating greater fear? If this formal analogy is accurate, then the $v_1 \rightarrow v_1$ and $v_2 \rightarrow v_2$ loops, say in Fig. 8c, would be rudimentary formal analogs of MFB pathways. In particular, MFB

pathways would be descending and ascending, as are the $v_1 \rightarrow v_1$ and $v_2 \rightarrow v_2$ loops [49]. This formal possibility is also compatible with the fact that NE is released by electrical stimulation of the MFB [19].

The formal "cholinergic system" of Fig. 8 encodes memory in $\mathscr{S} \rightarrow \mathscr{A}_f$ synapses. Many workers have reported that anticholinergics disrupt memory in punishment and avoidance situations [4, 7, 10, 18, 33, 39, 43, 46, 57, 61a]. Some details of these data will be cited and compared with formal network properties.

Operant behavior that is suppressed by punishment is disinhibited by cholinergic blockade of the VMH [44, 54]. Is this formally analogous to inhibition of $\mathscr{A}_f^+$ response to $\mathscr{S} \rightarrow \mathscr{A}_f^+$ channels and thus to a reduction of inhibitory $\mathscr{A}_f^+ \rightarrow \mathscr{S}$ feedback? Hypothalamic self-stimulation involves a cholinergic process, since rats learn to press a lever to self-inject carbachol (an ACh mimicker) into the LH [50]. Denote the lever-press response by R. Then the Olds *et al.* data have the following possible formal interpretation. Suppose carbachol injection is analogous to exciting v_2 in Fig. 8. Then carbachol acts as an input source to v_2 much as J does at v_1. This input brings $\mathscr{S}(R) \rightarrow \mathscr{A}_f^-$ sampling into the suprathreshold range and releases $\mathscr{A}_f^- \rightarrow \mathscr{S}$ feedback which drives $\mathscr{S}(R) \rightarrow \mathscr{M}$ sampling of the motor controls of R. Thus formal "carbachol injection" at v_2 acts much like an electrical self-stimulation pulse at v_2 [27]. This interpretation is strengthened by the fact that, *in vivo*, carbachol in LH facilitates the CAR, whereas cholinergic blockage by atropine or local depression by pentobarbital impairs the CAR [53]. Also compatible is the fact that rewarding hypothalamic stimulation reduces the aversive properties of shock [15]. This latter effect is formally analogous to inhibiting $\mathscr{A}_f^+$ response by increasing $\mathscr{A}_f^-$ input. Sepinwall [53] also reports the paradoxical result that carbachol in LH *and* in VMH facilitates the CAR. The VMH facilitation is not so large as the LH facilitation. By contrast, atropine in VMH reduces response suppression and in LH interferes with the CAR. This asymmetry in the effects of carbachol and atropine at LH and VMH would be understandable if carbachol in VMH were helping to drive a rebound in LH. This would be analogous to the following situation. Let feedback cues corresponding to nonavoidance responses during shock activate the sensory representations $\mathscr{S}$(non AR). The $\mathscr{S}$(non AR)s are conditioned to $\mathscr{A}_f^+$, since they are active when shock is on. Suppose that when the avoidance response occurs, the $\mathscr{S}$(non AR)s are shut off, and the feedback cues of the avoidance response activate the sensory representations $\mathscr{S}$(AR). In other words, we suppose that the cues of the avoidance response and of nonavoidance responses are substantially disjoint. Offset of $\mathscr{S}$(non AR)s removes input to $\mathscr{A}_f^+$, which thereupon drives a rebound at $\mathscr{A}_f^-$ that is sampled by $\mathscr{S}$(AR)s. This interpretation of CAR facilitation by carbachol in VMH can

be experimentally tested. The effect should be greatest in experiments where the overlap of $\mathscr{S}$(non AR) and $\mathscr{S}$(AR) cues is minimal. Then if carbachol is present in VMH during $\mathscr{S}(\text{non AR}) \rightarrow \mathscr{A}_f^+$ conditioning, it will strengthen the $\mathscr{S}(\text{non AR}) \rightarrow \mathscr{A}_f^+$ channels. The avoidance response, by shutting off the relatively large $\mathscr{A}_f^+$ input from $\mathscr{S}$(non AR)s, then yields a relatively large rebound at $\mathscr{A}_f^-$, which is learned by $\mathscr{S}$(AR)s. Carbachol in VMH has thereby enhanced CAR conditioning by indirectly strengthening $\mathscr{S}(\text{AR}) \rightarrow \mathscr{A}_f^-$ channels. Carbachol in LH also enhances CAR conditioning by directly strengthening $\mathscr{S}(\text{AR}) \rightarrow \mathscr{A}_f^-$ channels. If, by contrast, the experiment merely calls for fear conditioning, then carbachol in VMH should enhance the rate of fear conditioning and subsequent response suppression [44].

Given this interpretation of carbachol action, the effect of atropine has the following interpretation. Atropine in VMH reduces response suppression by reducing the conditioned $\mathscr{S} \rightarrow \mathscr{A}_f^+$ input by blocking $\mathscr{S} \rightarrow \mathscr{A}_f^+$ synapses. Atropine cannot drive a rebound at $\mathscr{A}_f^-$ because it reduces the likelihood that termination of $\mathscr{S} \rightarrow \mathscr{A}_f^+$ inputs can drive a significant rebound at $\mathscr{A}_f^-$. Atropine in LH can nonetheless interfere with the CAR by reducing the positive incentive motivation controlled by $\mathscr{S} \rightarrow \mathscr{A}_f^-$ channels.

Carlton ([11; see also Khavari [40]) argues that anticholinergics act by inhibiting habituation of incorrect responses, rather than by disrupting memory of correct responses. Some of his remarks in special support of the habituation hypothesis are compatible with memory disruption effects on Fig. 8; for example, concerning the influence of carbachol and anticholinergics on VMH [11, page 324]. Carlton [11, page 305] also argues against memory disruption by showing that scopolamine (an ACH inhibitor) produces more profound deficits with negative than with positive rewards. This argument should be supplemented by dose-dependent studies, since positive and negative reward systems have different thresholds in our networks: a shock input by itself can drive a CER; food input must summate with the prevailing internal hunger input before it can release $\mathscr{S} \rightarrow \mathscr{M}$ signals. Carlton [11, page 323] suggests that the VMH contains ACh and that the LH contains NE. Our present interpretation finds ACh and NE at parallel loci in both VMH and LH, subject to the qualification that interneurons, such as those comprising the $v_1 \rightarrow v_2$ and $v_2 \rightarrow v_2$ loops in Fig. 8c, might interconnect other nuclei in parallel with LH and VMH, and thereby separate the two transmitter systems. The present interpretation also differs from the view [40, 57] that the cholinergic system mediates punishment and the adrenergic system mediates reward.

10. REGULATION OF TOTAL $\mathcal{S} \rightarrow \mathcal{A}$ AND TOTAL $\mathcal{A} \rightarrow \mathcal{S}$ INPUTS BY NONSPECIFIC INHIBITORY INTERNEURONS

Section 6 notes that the total $\mathcal{S} \rightarrow \mathcal{A}$ input must be regulated to keep the total I input away from the extremes in the inverted U. Another reason for doing this is the following. CSs can differ in sensory complexity and intensity. Thus variable numbers of $\mathcal{S}$ representations can be excited at different times, and each $\mathcal{S}$ representation can be excited to a different degree. Suppose that the representations send signals to $\mathcal{A}$ that are independent of each other. Consider two different CSs, CS_1 and CS_2. Let CS_1 excite the N $\mathcal{S}$ representations $\mathcal{S}_1$, and CS_2 excite the MN $\mathcal{S}$ representations $\mathcal{S}_2$, each to the same degree D. Suppose that CS_1 can be learned as the cue for a lever-press response for food. On recall trials, conditioned $\mathcal{S}_1 \rightarrow \mathcal{A}$ signals can therefore release $\mathcal{A} \rightarrow \mathcal{S}$ feedback *if* $\mathcal{O}$ is hungry; $\mathcal{S}_1 \rightarrow \mathcal{M}$ signals then drive the lever press. CS_2 can drive learning and performance of the lever press even if $\mathcal{O}$ is not hungry, if M is sufficiently large. This is because total $\mathcal{S} \rightarrow \mathcal{A}$ input increases with M until $\mathcal{S} \rightarrow \mathcal{A}$ input alone can overcome the $\mathcal{A} \rightarrow \mathcal{S}$ spiking threshold. In a similar fashion, CS_1 can drive the lever press without hunger if D is sufficiently large. To avoid these catastrophes, $\mathcal{S} \rightarrow \mathcal{A}$ signals cannot be independently delivered; the total $\mathcal{S} \rightarrow \mathcal{A}$ input must be regulated.

There exists at least two ways to do this. The general mechanism is known as *pattern normalization*, and was introduced in [25]. Two possibilities are illustrated in Fig. 9: Figure 9a illustrates a subtractive, nonrecurrent, nonspecific, inhibitory interneuron, or "horizontal cell" [25, Sections 8–10]. Such interneurons can truncate the total $\mathcal{S} \rightarrow \mathcal{A}$ input when it reaches a fixed maximum. In the simplest case, let the output of $\mathcal{S}_i$ be $I_i(t)$ and the net output after operation of an inhibitory interneuron at layer $\tilde{\mathcal{S}}$ be $f_i(t) = I_i(t) - [\Sigma_{k=1}^{n} I_k(t) - \Gamma]^+$, with all $I_i(t) \geq 0$

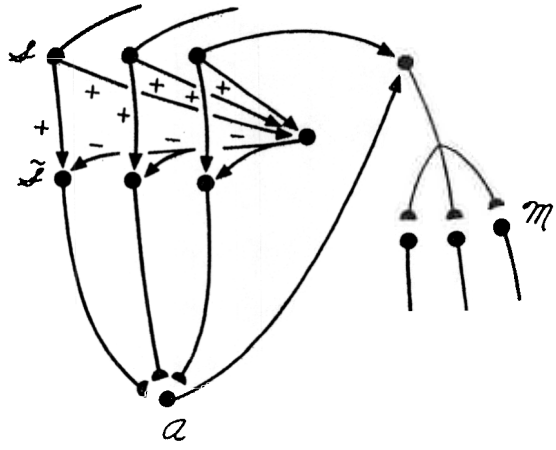

FIG. 9a. $\mathcal{S} \rightarrow \mathcal{A}$ Normalization: subtractive inhibition using nonspecific interneurons.

and $\Gamma > 0$. Clearly $f_i = I_i$ whenever $\Sigma_{k=1}^n I_k \leqslant \Gamma$. Thus $f_i = I_i \leqslant \Sigma_{k=1}^n I_k \leqslant \Gamma$ in this case. If $\Sigma_{k=1}^n I_k \geqslant \Gamma$, then $\Sigma_{k=1}^n f_k = (1 - n)\Sigma_{k=1}^n I_k + n\Gamma \leqslant \Gamma$. In all cases, the total output $\Sigma_{k=1}^n f_k$ is bounded by Γ.

Figure 9b illustrates an on-center off-surround field with shunting excitation and inhibition [25, Section 14A]. In the simplest case, let the ith cell in the $\tilde{\mathscr{S}}$ layer of Fig. 9b have potential

$$\dot{x}_i = (M - x_i)I_i - \alpha x_i - x_i \sum_{k \neq i} I_k.$$

This is a passive membrane equation, with equilibrium scaled at zero for convenience, and inputs I_i representing depolarizing or hyperpolarizing conductance changes. It can be shown that the total output from $\tilde{\mathscr{S}}$ to $\mathscr{A}$ is bounded by M, and that each x_i is asymptotically proportional to the pattern weight $\theta_i = I_i(\Sigma_{k=1}^n I_k)^{-1}$. One can also study influences of different thresholds, time lags, exponential averaging rates, and axonal path weights in excitatory and inhibitory cells, variations in total output due to variations in input pattern, and so on. Each of these normalization mechanisms has particular advantages, which Grossberg [25] studies.

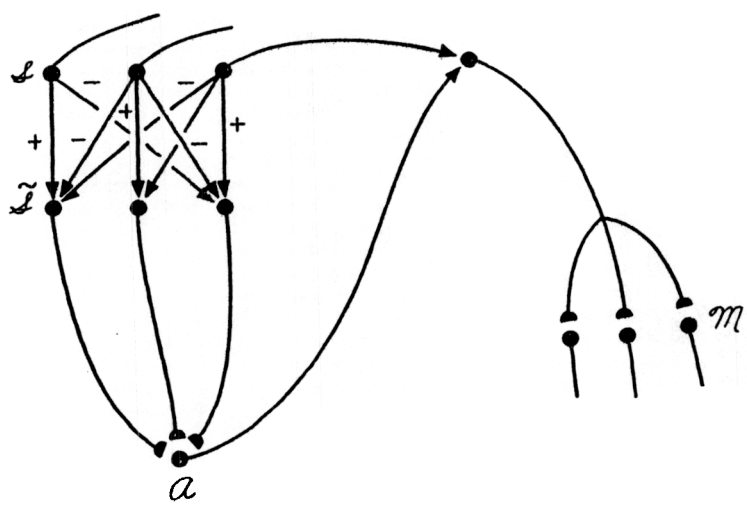

FIG. 9b. $\mathscr{S} \to \mathscr{A}$ Normalization: shunting inhibition using on-center off-surround field.

These normalization mechanisms form part of the filtering mechanism that permits only prescribed stimulus features, or classes of patterns, to excite particular $\mathscr{S}$ representations. Thus it is possible that the filtering mechanism, by creating selective $\mathscr{S}$ channels, automatically regulates the total $\mathscr{S} \to \mathscr{A}$ output.

The total $\mathscr{A} \to \mathscr{S}$ feedback input must also be regulated to prevent this input from indiscriminately firing $\mathscr{S} \to \mathscr{M}$ channels in the absence of sensory cues to these channels. Inhibitory interneurons therefore modify the outputs of the various arousal sources before they reach $\mathscr{S}$. The next

section studies some anatomies that incorporate inhibitory interneurons that achieve this goal by satisfying another basic principle of network design.

11. A SENSORY-DRIVE HETERARCHY

Do the inhibitory interneurons that regulate total $\mathscr{A} \rightarrow \mathscr{S}$ feedback operate before or after the stage at which $\mathscr{S} \rightarrow \mathscr{A}$ signals combine with internal drive inputs? The answer is "after" if we accept the next postulate.

Postulate XIII. $\mathcal{O}$ can (sometimes) consummate drive D_1, even when drive D_2 is higher, if sensory cues appropriate to D_1 are available whereas cues appropriate to D_2 are not available.

For example, many $\mathcal{O}$s can eat if food is regularly available, even if their sex drives become very high in the absence of a mate. Consider Fig. 10: In Fig. 10a, the internal homeostatic inputs representing different

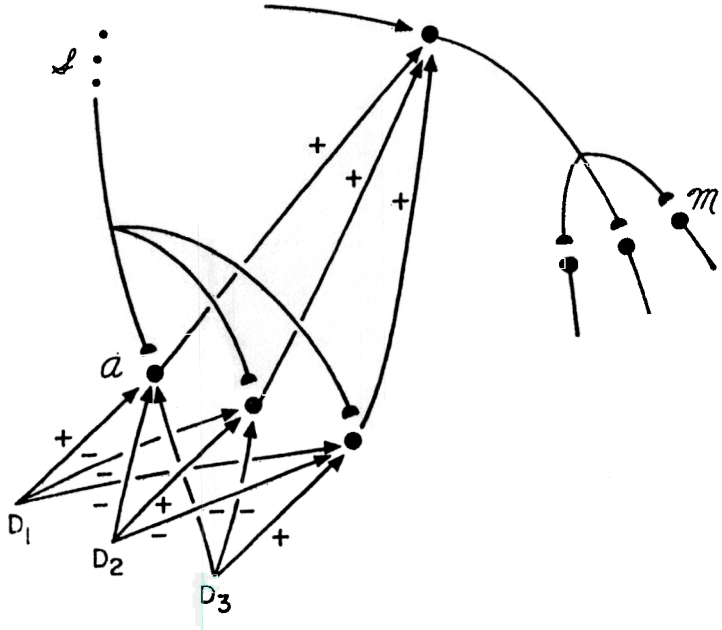

FIG. 10a. Only prepotent drive can release behavior.

drives inhibit each other before $\mathscr{S}$ can affect them. Only one drive representation receives a net positive input after operation of the nonspecific inhibitory interneurons. Only this drive can be a source for $\mathscr{A} \rightarrow \mathscr{S}$ feedback and motor output. If the $\mathscr{S}$ cues needed to release this feedback are not available, $\mathcal{O}$ will not satisfy any drive. This $\mathcal{O}$ will starve in the absence of sex. In Fig. 10b, sensorily driven $\mathscr{S} \rightarrow \mathscr{A}$ inputs summate with internal homeostatic inputs before the inhibitory interneurons operate. Thus a positive, but not prepotent, drive can release $\mathscr{A} \rightarrow \mathscr{S}$ feedback and compatible motor output if sensory cues appropriate to this drive predominate.

This $\mathcal{O}$ can eat and wait for sex. This sensory-drive heterarchy seems related to data of Cox and Valenstein [16], who show that different sensory cues can release different behavior in the presence of hypothalamic stimulation at a fixed spatial locus. Analogous data are collected by Kopa *et al.* [41], who stimulated an area dorsal to the centrum medianum nucleus of the thalamus.

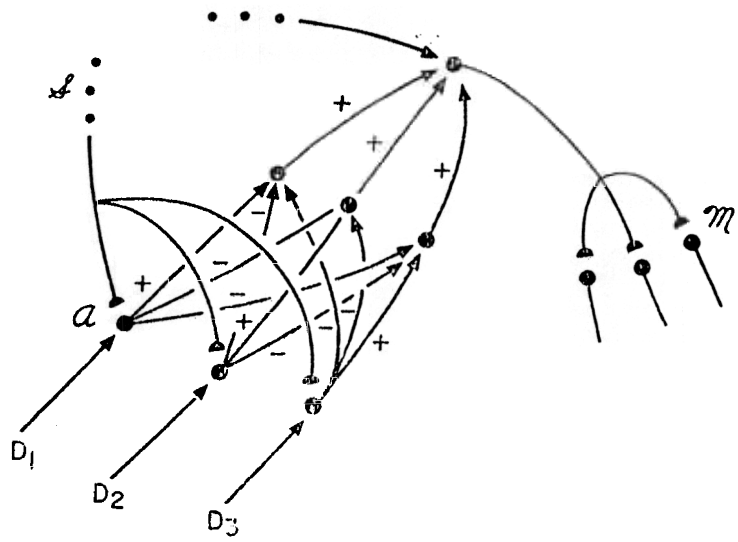

FIG. 10b. Mixing of sensory and drive cues in heterarchical anatomy.

The relative importance of $\mathscr{S} \to \mathscr{A}$ versus internal homeostatic inputs can be explicitly computed in specific cases. Note that the $\mathscr{S} \to \mathscr{A}$ input received at one drive representation is influenced by the pattern of $\mathscr{S} \to \mathscr{A}$ inputs sent to *all other* drive representations, due to the action of nonspecific inhibitory interneurons. Similarly, the distribution of $\mathscr{A} \to \mathscr{S}$ feedback is influenced by the pattern of all sensory-plus-drive combinations. This is a highly nonlocal system.

New sources of pathological $\mathscr{A} \to \mathscr{S}$ overarousal are now evident; for example, an increase in the threshold Γ of nonspecific inhibitory interneurons, as in Fig. 9a; or an increase in the saturation level M of potentials in cells of layer $\mathscr{S}$, as in Fig. 9b. The parametric studies in Grossberg [25] reveal still other possible sources of overarousal.

The work reported in this article was supported in part by the Alfred P. Sloan Foundation and the Office of Naval Research (N00014–67–A–0204–0051).

REFERENCES

1 N. E. Andén, A. Carlsson and J. Häggendal, Adrenergic Mechanisms, *Ann. Rev. Pharmacol.* (H. W. Elliott, W. C. Cutting and R. H. Dreisbach eds.) **9**, 119–134 (1969).

2 Z. Annau and L. J. Kamin, The conditioned emotional response as a function of intensity of the US, *J. Comp. Physiol. Psychol.* **54**, 428–432 (1961).

3 A. B. Ashton, S. C. Bitgood and J. W. Moore, Auditory differential conditioning of the rabbit nictitating membrane response: III. Effects of US shock intensity and duration, *Psychonomic Sci.* **15,** 127–128 (1969).
4 B. Berger, D. L. Margules and L. Stein, Prevention of learning of fear by oxazepam and scopolamine. *Amer. Psychol.* **22,** 492 (1967).
5 F. E. Bloom and N. J. Giarman, Physiological and Pharmacological Considerations of Biogenic Amines in the Nervous System, *Ann. Rev. Pharmacol.* (H. W. Elliott, W. C. Cutting and R. H. Dreisbach, eds.) **8,** 229–258 (1968).
6 E. E. Boe, Effect of punishment duration and intensity on the extinction of an instrumental response. *J. Exptl. Psychol.* **72,** 125–131 (1966).
7 Z. Bohdanecky and M. E. Jarvik, Impairment of one-trial passive avoidance learning in mice by scopolamine, scopolamine methylbromide, and physostigmine, *Intern. J. Neuropharmacol.* **6,** 217–222 (1967).
8 J. J. Boren, M. Sidman and R. J. Herrnstein, Avoidance, escape and extinction as functions of shock intensity. *J. Comp. Physiol. Psychol.* **52,** 420–425 (1959).
9 G. Borozci, L. H. Storms and W. E. Broen, Response suppression and recovery of responding at different deprivation levels as functions of intensity and duration of punishment. *J. Comp. Physiol. Psychol.* **58,** 456–459 (1964).
10 O. Burešova, J. Bureš, Z. Bohdanecky and T. Weiss, Effect of atropine on learning, extinction, retention and retrieval in rats. *Psychopharmacologia* **5,** 255–263 (1964).
11 P. L. Carlton, Brain-acetylcholine and inhibition, in *Reinforcement and Behavior* (J. T. Tapp, ed.), Academic Press, New York (1969).
12 B. A. Campbell, Interaction of aversive stimuli: Summation on inhibition? *J. Exptl. Psychol.* **78,** 181–190 (1968).
13 R. M. Church, Response suppression, in *Punishment and Aversive Behavior* (B. A. Campbell and R. M. Church, eds.), Appleton, New York (1969).
14 R. M. Church, G. A. Raymond and R. D. Beauchamp, Response suppression as a function of intensity and duration of punishment. *J. Comp. Physiol. Psychol.* **63,** 39–44 (1967).
15 V. C. Cox and E. S. Valenstein, Attenuation of aversive properties of peripheral shock by hypothalamic stimulation. *Science* **149,** 323–325 (1965).
16 V. C. Cox and E. S. Valenstein, Effects of stimulus intensity on behavior elicited by hypothalamic stimulation, *J. Comp. Physiol. Psychol.* **69,** 730–733 (1969).
17 M. R. D'Amato, J. Fazzaro and M. Etkin, Discriminated bar-press avoidance maintenance and extinction in rats as a function of shock intensity. *J. Comp. Physiol. Psychol.* **63,** 351–354.
18 S. S. Fox, Self-maintained sensory input and sensory deprivation in monkeys: A behavioral and neuropharmacological study, *J. Comp. Physiol. Psychol.* **55,** 438–444 (1962).
19 K. Fuxe, T. Hökfelt, and U. Ungerstedt, Morphological and functional Aspects of Central Monoamine Neurons, in *Intern. Rev. Neurobiol.* (C. C. Pfeiffer and J. R. Symythies, eds.) **13,** 93–126. Academic Press, New York (1970).
20 W. J. Gardner, J. C. R. Licklider and A. Z. Weisz, Suppression of pain by sound. *Science* **132,** 32–33 (1961).
21 S. P. Grossman, The VMH: a center for affective reactions, satiety, or both? *Physiol. Behavior* **1,** 1–10 (1966).
22 S. Grossberg, Embedding fields: A theory of learning with physiological implications. *J. Math. Psychol.* **6,** 209–239 (1969).
23 S. Grossberg, On learning, information, lateral inhibition, and transmitters, *Math. Biosci.* **4,** 255–310 (1969).

24 S. Grossberg, On the production and release of chemical transmitters and related topics in cellular control, *J. Theoret. Biol.* **22,** 325–364 (1969).
25 S. Grossberg, Neural pattern discrimination. *J. Theoret. Biol.* **27,** 291–337 (1970).
26 S. Grossberg, Embedding Fields: Underlying philosophy, mathematics, and applications to psychology, physiology, and anatomy, *J. Cybernet.* **1,** 28–50 (1971).
27 S. Grossberg, On the dynamics of operant conditioning. *J. Theoret. Biol.* **33,** 225–255 (1971).
28 S. Grossberg, Neural expectation: Cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes, *Kybernetik* (1972).
29 S. Grossberg, A neural theory of punishment and avoidance, I, Qualitative theory, *Math. Biosci* **15,** 39–68 (1972).
30 S. Grossberg and J. Pepe, Schizophrenia: Possible dependence of associational span, bowing, and primacy vs. recency on spiking threshold, *Behav. Sci.* **15,** 359–362 (1970).
31 S. Grossberg and J. Pepe, Spiking threshold and overarousal effects in serial learning, *J. Statist. Phys.* **1,** 319–350 (1971).
32 J. A. Harvey and C. E. Lints, Lesions in the medial forebrain bundle: Delayed effects on sensitivity to electric shock, *Science* **148,** 250–251 (1965).
33 R. H. Haude, Effects of scopolamine on discriminated responding in the rat. *J. Pharmacol. Exp. Therap.* **126,** 349–358 (1959).
34 D. O. Hebb, Drives and the CNS (conceptual nervous system), *Psychol. Rev.* **62,** 243–254 (1955).
35 F. W. Huff, T. P. Piantanida and G. L. Morris, Free operant avoidance responding as a function of serially presented variations in UCS intensity, *Psychonomic Sci.* **8,** 111–112 (1967).
36 J. L. Johnson and R. M. Church, Effects of shock intensity on nondiscriminative avoidance learning of rats in a shuttlebox, *Psychonomic Sci.*, **3,** 497–498 (1965).
37 L. J. Kamin, C. J. Brimer and A. H. Black, Conditioned suppression as a monitor of fear in the course of avoidance training. *J. Comp. Physiol. Psychol.* **56,** 497–501 (1963).
38 J. D. Keehn, Effect of shock duration on Sidman avoidance response rates, *Psychol. Repts*, **13,** 852 (1963).
39 K. A. Khavari, Effects of central versus intraperitoneal d-amphetamine administration on learned behavior, *J. Comp. Physiol. Psychol.* **68,** 226–234 (1969).
40 K. A. Khavari, Adrenergic-cholinergic involvement in modulation of learned behavior, *J. Comp. Physiol. Psychol.* **74,** 284–291 (1971).
41 J. Kopa, I. Szabó and E. Grastyán, A dual behavioral effect from stimulating the same thalamic point with identical stimulus parameters in different conditional reflex situations, *Acta Physiol. Acad. Sci. Hungaricae (Budapest)*, **21,** 207–214 (1962).
42 C. Kornetsky and M. Eliasson, Reticular stimulation and chlorpromazine: an animal model for schizophrenic overarousal, *Science* **165,** 1273–1274 (1969).
43 D. Macht, A pharmacodynamic analysis of the cerebral effects of atropin, homatropin, scopolamin and related drugs, *J. Pharmacol. Exptl. Therap.* **22,** 35–48 (1923).
44 D. L. Margules and L. Stein, Neuroleptics vs. tranquillizers: Evidence from animal behavior studies of mode and site of action, in *Neuropsychopharmacology* (H. Brill, ed.), International Congress Series No. 129. Excerpta Medica Foundation, Amsterdam (1967).
45 L. K. Martin and D. Reiss, Effects of US intensity during previous discrete delay conditioning on conditioned acceleration during avoidance conditioning, *J. Comp. Physiol. Psychol.* **69,** 196–200 (1969).

46 W. R. Miles, Hyoscine vs. alcohol: Effect on the behavior of rats in the elevated maze, *Proc. 9th Intern. Congr. Psychol.* (1929), pp. 309–310.
47 N. E. Miller, Learning resistance to pain and fear: Effects of overlearning, exposure, and rewarded exposure in context, *J. Exptl. Psychol.* **60,** 137–145 (1960).
48 A. K. Myers, Effects of continuous loud noise during instrumental shock-escape conditioning, *J. Comp. Physiol. Psychol.* **68,** 617–622 (1969).
49 W. J. H. Nauta and W. Haymaker, Hypothalamic nuclei and fiber connections, in *The Hypothalamus* (W. Haymaker, E. Anderson and W. J. H. Nauta, eds.), C. C. Thomas, Springfield, Illinois (1969).
50 J. Olds, A. Yuwiler, M. E. Olds and C. Yun, Neurohumors in hypothalamic substates of reward, *Amer. J. Physiol.* **207,** 242–254 (1964).
51 M. I. Phillips and P. B. Bradley, Reticular stimulation and chlorpromazine, *Science* **168,** 1122–1123 (1970).
52 D. Reiss, Sidman avoidance in rats as a function of shock intensity and duration, *J. Comp. Physiol. Psychol.* **73,** 481–485 (1970).
53 J. Sepinwall, Enhancement and impairment of avoidance behavior by chemical stimulation of the hypothalamus, *J. Comp. Physiol. Psychol.* **68,** 393–399 (1969).
54 J. Sepinwall and F. S. Grodsky, Effects of cholinergic stimulation of the rat hypothalamus or cholinergic blockade on discrete-trial conflict. Paper presented at the meeting of the Eastern Psychological Association, Washington, April (1968).
55 M. H. Sheard, J. B. Appel, D. X. Freedman, *J. Psychiat. Res.* **5,** 237–242 (1967).
56 C. D. Shute and P. R. Lewis, Cholinergic and monoaminergic pathways in the hypothalamus, *Brit. Med. Bull.* **22,** 221–226 (1966).
57 L. Stein, Noradrenergic substrates of positive reinforcements: Site of motivational action of amphetamine and chlorpromazine, in *Neuropsychopharmacology* (H. Brill *et al.*, eds.), Excerpta Med. Foundation, Amsterdam (1967).
58 A. Strouthes, Effect of CS-onset, UCS-termination delay, UCS duration, CS-onset interval, and number of CS-UCS pairings on conditioned fear response. *J. Exptl. Psychol.* **69,** 287–291 (1965).
59 R. F. Thompson, *Foundations of Physiological Psychology*, Harper, New York (1967).
60 N. Weiner, Regulation of norepinephrine biosynthesis, in *Ann. Rev. Pharmacol.* (N. W. Elliott, W. C. Cutting, and R. H. Dreisbach, eds.) **10,** 273–290 (1970).
61 B. Weiss and V. G. Laties, Behavioral Pharmacology and Toxicology, in *Ann. Rev. Pharmacol.* (H. W. Elliott, W. C. Cutting and R. H. Dreisbach, eds.), **9,** 297–326 (1969).
62 C. Wise and L. Stein, Facilitation of brain self-stimulation by central administration of norepinephrine, *Science* **163,** 299–301 (1969).