THEORY OF THE HUMAN OPERATOR IN CONTROL SYSTEMS[1]

I. THE OPERATOR AS AN ENGINEERING SYSTEM

By the late KENNETH J. W. CRAIK

Medical Research Council, Applied Psychology Unit, University of Cambridge

I. *The human operator behaves basically as an intermittent correction servo* (p. 56). II. *The intermittent corrections consist of 'ballistic' movements* (pp. 56–58). III. *There are some counteracting processes tending to make controls seem continuous* (pp. 58–59). IV. *Electrical models could fairly exactly simulate the human operator's behaviour in tracking* (pp. 59–61).

I. *The human operator behaves basically as an intermittent correction servo.* The evidence for this is the periodic or 'wavy' nature of the time-record of tracking errors, showing a spectrum with a predominant frequency of about 0·5 sec. with a smaller cluster of frequencies from 0·25 to 1 sec. This periodicity might be attributed to a sensory threshold or 'dead zone', such that misalinements smaller than a certain value evoke no corrective movement; but there is evidence against this. First, the display-magnification is usually such that the misalinements occurring during steady tracking exceed the known threshold (i.e. visual acuity). Secondly, if the rate of the course, or the magnification of the display for a given course, is increased by a certain factor, the periodicity of the corrections is little, if at all, affected; whereas if their periodicity were determined by the time taken for the misalinements to reach a certain 'threshold' value this alteration should shorten the periodicity of the corrections in the same ratio.

Consistently with the above principle, we find that the mean error in tracking any given variable-direction course is nearly proportional to the rate of the course, over a wide range of speeds. We should account for this by saying that the faster the course the greater the misalinements that occur in each period between two corrections, in strict proportion.

II. *The intermittent corrections consist of 'ballistic' movements.* For example, they have a predetermined time-pattern and are 'triggered off' as a whole. This behaviour may be contrasted with that of an intermittent correction servo in which, for instance, a follow-up motor is intermittently switched into a circuit in which it runs until it has reduced the misalinement to zero, and this action reduces the input to it to zero so that the motor stops. In the human operator, on the other hand, at a particular instant (i.e. about 0·3 sec.

[1] *Editor's note.* The material of this paper and of a second one to be published in the next number of this *Journal*, was prepared by the late Dr Craik in March and April 1945. It was written substantially in the form in which it now appears, and undoubtedly was intended, not as a final form for publication, but as a preliminary draft for discussion. In view, however, of the very lively and widespread interest that has developed in the problems discussed in these two papers, and of the attractive and original way in which Craik handles them, it seems highly desirable that they should be made public. The present paper deals with the possibility and significance of the construction of mechanical and electrical models for simulating human behaviour in any kind of continued pursuit task. The second paper deals with fundamental physiological and psychological mechanisms which appear to be involved in operations of the same kind.

Most of the work of preparing the articles for publication was done by Miss M. A. Vince.

after the end of the preceding corrective movement), a corrective movement having a predetermined time-course (usually occupying about 0·2 sec.) is triggered off. The evidence for this is based on studies of reaction time, i.e. on the internal time-lag of the operator, due to the time taken by the sense-organ to respond, for the nerve-impulses to traverse the central nervous system, for the appropriate response to be 'selected' and for the nerve impulses reaching it to traverse the motor nerves. This lag is about 0·2–0·3 sec. Thus, if a human operator's limb movement amplitude, or velocity, or acceleration, were determined continuously by the misalinement, continued oscillations of approximately 0·5 sec. period, and of whatever amplitude they commenced at, would inevitably result. This tendency could be overcome if a misalinement triggered off a ballistic movement of fairly correct amplitude (say $\pm 10\%$); the eye would then detect the residual misalinement, which may be composed of two parts—the error in the first ballistic movement, and any movement of the target which has occurred in the meantime. In the absence of the latter (e.g. in aiming at a stationary target) the second corrective movement may again be accurate to $\pm 10\%$ of its own value so that the misalinement is reduced to 1 % of its original value.

Direct evidence bearing on this ballistic behaviour can be obtained in various ways. First it is easy to present a misalinement to an operator, and then screen his eyes just before he makes his corrective movement. A movement accurate to within about 10 % will result, if he has previously learned the 'feel' of the control. If he operated like a follower motor, intermittently switched in, such obscuration of the misalinement would be equivalent to cutting the input connexions, and the motor would not of course make any further corrective movement.

Physiologists might make a further hypothesis to avoid the 'ballistic' theory. They might say that the visual misalinement, once it has been detected by the eye, becomes translated into a 'limb-movement-misalinement' (i.e. a kinaesthetic misalinement) which acts as the continuous input to the limb until our kinaesthetic (i.e. joint and muscle senses) register 'correct position', or no misalinement. We certainly do possess such a sense, but it is less easy to abolish it experimentally and to see what happens in its absence, than in the case of vision. Patients with *tabes dorsalis* have considerable loss of kinaesthetic sensation, but it is difficult to know how complete this loss is in any particular case.

In any case, the same general argument as before—that a continuous series of oscillations of the initial amplitude would occur if elimination of kinaesthetic misalinement were the sole determinants of movements, owing to the inevitable reaction-time lag—seems to apply. Further, it is possible to show that considerable precision of movement is maintained even when the movements are made so rapidly that they are completed before the kinaesthetic stimulus corresponding to their first approximation to the right position could have 'gone the round' of the central nervous system and controlled the subsequent output-movement; unless indeed reaction-times to kinaesthetic stimuli were vastly shorter than to any other kind. It is, however, possible to show experimentally that kinaesthetic reaction times are very little shorter than auditory, for instance. Thus, we may ask the operator to move a lever against a stiff spring, so as to make a rapid movement to correct a misalinement, and then to return to his starting point. After he has learned the 'feel' of the control (i.e. its gear ratio and spring tension) and is making fairly accurate movements, we suddenly alter the spring tension, so that he over- or under-shoots. A record of this on a fast-moving drum shows that about 0·15 sec. elapses after he begins his movement, until he is able to begin a readjustment of it, to meet the modification of resistance.

Similarly, it is possible to show that in playing musical instruments, typewriting, sending morse, etc., complicated patterns of movement are executed at a rate which would be impossible if they were continuously governed by the value of the misalinement, with the inevitable reaction-time lag. Apparently they must be individually performed, triggered off ballistically, and the sensory feed-back must take the form of a delayed modification of the amplitude of subsequent movements. Sensory control, in other words, alters the 'internal gear ratio' or amplification of the operator with a time lag and determines whether subsequent corrective movements will be made; it does not govern the amplitude of each individual movement while it is being made. We could make a servo, using existing engineering principles, which would show the features of intermittent, ballistic correction. But this last point—the fact that the sensory misalinement alters not merely the amplitude of the response but the *relation* between the input, or misalinement and the output, or response—introduces a further complication, the nearest approach to which, in engineering, seems to be 'floating plus proportional control'. Even this involves a quantitative alteration of the amplification of the system by the residual misalinement, whereas something more complex still seems to be occurring—a very wide alteration in the functional relationship between input and output.

For instance, if the operator is using a positional control, his successive ballistic corrective movements should be linearly proportional to the misalinements; but if he is using a velocity control they would have to be linearly proportional to the *derivatives* of the misalinements. Roughly speaking, we might call this *qualitative* modification or output on a basis of some response to the difference between instantaneous input and output at the previous instant, or 'qualitative feed-back'.

A further complication is introduced by the operator's ability to 'anticipate' movements of the target, or alterations in misalinement. For instance, with a positional control, the errors in tracking a moving target are usually less than we should predict on the above theory of intermittent ballistic corrections. The operator goes on turning the handle steadily, or even accelerating it; his record, after some practice, becomes much smoother than it was initially; and if he finds that he is still lagging behind the target (as if the above theory is correct he is bound to do) he can put in an extra forward movement. Here we have several processes.

III. *There are some counteracting processes tending to make controls seem continuous.* First, there is one akin to momentum, or inertia. If the operator has been turning the handle, in a series of discrete movements, for some seconds, he will tend to convert this into as steady a rotation as he is capable of, and to continue doing so although the misalinement may be zero, i.e. he has zero input to produce this output! (This can be shown experimentally by suddenly stopping the target, when the operator will overshoot for the period of one reaction time, until the serious misalinement which results stops his steady output from continuing.) It is for this reason that in Section I it was stated that the human operator is *basically* an intermittent correction servo; he has in addition this mechanism for going on doing whatever is giving a satisfactory result, or zero misalinement, rather like a heavy flywheel, and having the same valuable smoothing effect.

What are the essential features of this process, and can we conceive any mechanisms which will accomplish it? When the operator continues to turn the handle at the same speed, independently of whether there is any input or not, he is, in humanistic terms, assuming that he is justified in doing so, in order to compensate for his reaction time lag. Since he is always subject to this lag, in attempting to keep up with the present he is always in fact being a prophet and extrapolating from past data! It is really no different from the further kind of anticipation which enables him to extrapolate into the physical future. Now all scientific prediction consists in discovering, in data of the distant past and of the immediate past which we incorrectly call the present, laws or formulae which apply also to the future, so that if we act in accordance with those laws our behaviour will be appropriate to that future when it becomes the present. Thus the essential feature of extrapolation and anticipation is, again in humanistic terms, that the operator should detect the *constants* in what he is doing. Thus, he may move a handwheel in a series of jerks, so that its *position* changes from moment to moment, but he may realize after a few seconds that he is turning it at a *steady rate*, i.e. its *angular velocity* is constant; and having discovered this he may try whether it will not pay him to go on doing so; usually it will. He may, however, find that its rate is changing—the target has angular acceleration. He may, in theory, at any rate, be able to feel this acceleration which he is having to put into the handwheel, and if he happens to find that it is constant, or nearly so, and is able to put out a constant acceleration of this value in turning the wheel, again he may achieve better following.

Now let us look at the same thing from a mechanical point of view. There are many devices—such as speedometers and accelerometers—which do the differentiations involved in recording velocities and accelerations; and the problem would be how, for instance, to couple a number of such devices to a telephone selector-switch operating motor, so that if the output of the motor over a few seconds of intermittent corrective action showed a constant reading on the speedometer, or even on the accelerometer, the motor would be caused to go on putting out this speed or acceleration, irrespective of whether there was any input or not, unless or until such behaviour gave rise to a large misalinement. If that happened the extrapolating system would be overridden and intermittent corrections would begin again, until a new value for a constant was found. The solution would seem to be to provide the motor with positive feed-back of such a kind that it continued to go on doing whatever it was doing at the moment—running steadily or accelerating uniformly. Such a system would need considerable smoothing and stabilization, otherwise any slight disturbance, such as a slight acceleration, would very rapidly be cumulative, and the machine would reach its maximum speed; but if the feed-back were delayed and smoothed, the system could be sufficiently stable and would not 'wander' too badly. This system would of course be combined with negative feed-back of the ordinary kind (viz. actuation by the difference between input and output quantities), so that if the positive feed-back led to overestimation of the velocity, or if the target started to decelerate, the motor would overshoot and this would introduce a positional misalinement, which would reverse the direction of mechanical control. This would alter the average value, for the last time-interval, to the positive feed-back system, which would therefore cease from perpetuating this velocity but would, when it had time to steady down, start putting in a new one.

IV. *Electrical models could fairly exactly simulate the human operator's behaviour in tracking.* In general terms, the extraction of the inputs for the positive feed-back network consists of successive differentiations, while the extrapolations on the basis of them consist of successive integrations. Let us consider in more detail some circuits by which this might be accomplished. Suppose the motor drives a generator across whose output terminals is a capacity in series with a high resistance, constituting a differentiating circuit with a time-lag or averaging effect, owing to the time-constant of the system. Then the generator voltage is proportional to the speed of the motor and the voltage across the resistance of the differentiating system is proportional to its acceleration; if necessary, higher derivatives can be obtained in the same way. The lag in the first differentiation can be obtained by putting a resistance and capacity in series across the generator output and taking the voltage off the capacity. The output voltage from this smoothing system is taken to the input of the amplifier supplying the motor fields, and should cause the motor to continue running at the mean speed at which it has been manually rotated for a sufficient time to cause the voltage delayed and smoothed across the generator to reach a steady value. The speed of running will of course wander slowly in time if the system also has ordinary velodyne negative feed-back for velocity control. If the manually applied speed was an accelerating one, the system will maintain a mean steady speed if it is supplied with one differentiating stage only (i.e. the generator with its delaying system). But if there is a second differentiating system, with a longer time-constant, it will register a manually imposed change of velocity over several periods of operation of the first differentiating system, i.e. an acceleration, and if this is integrated by a resistance-capacity circuit and applied to the amplifier serving the fields, a uniform acceleration will occur.

Of course, it is not necessary for the original speeds to be put in manually; with a velodyne fitted up as a servo auto-following system in which the task of the velodyne is to make a slider keep on the centre of a potentiometer, for instance, which is moved by an external agency, the mechanical control will commence by ordinary positional following, being actuated by the misalinements. Further, though this alone would lead to a lag behind a uniformly moving target, once it has started to run at a constant velocity, the remaining misalinement will still be operative, if arranged to be in series with the positive feed-back voltage, so as to cause the shaft to step on and make up for the lag. This system would have many of the same characteristics as a velodyne with phase advance produced by delayed negative feed-back—i.e. a condenser across the generator output.

It should be possible to make a velodyne simulate the 'intermittent ballistic correction' process considered in principles (I) and (II). Thus, the error-voltage representing the misalinement could be connected periodically by a rotating contact to a condenser which is charged. This condenser would then be switched on to the amplifier input and would result in a 'ballistic' rotation of the output shaft through an angle proportional to the charge on the condenser.

Little is known of possible physiological mechanisms for accomplishing this kind of thing. There is evidence (e.g. from sensory adaptation and accommodation of nerve) of differentiating systems, at least of the first order, which may serve to measure rates of change of stimulation, though our knowledge extends only to stimulus *intensity* and not to more complex stimuli such as misalinements in space. Even here, it is possible to suggest hypothetical spatial differentiating systems which are not physiologically inconceivable. The other aspect—the integration, resulting from positive feed-back—would seem to require 'autorhythmic' nervous centres which continue to discharge once they have been forced to do so, and in a way which follows the original forcing stimulus. The beating of the isolated frog's heart and the spontaneous oscillatory potentials in the excised frog's brain and in the intact cortex of man (both Berger rhythms and the abnormal rhythms of epilepsy) are suggestive in this respect, for they are evidence of self-maintaining neural oscillators. Lorente de No's and Ransom's concept of the 'closed neurone circuit' would serve the same purpose. What has to be considered is clearly a form of positive feed-back and the main difficulty in all the cases just mentioned would seem to be that what is required is continuous feed-back of excitation in the form of nerve-impulses following after the neurones have recovered from their refractory phase, whereas slow potential oscillations probably imply discharges of some other kind than trains of nerve impulses.

We should also consider long-lasting changes of stimulus-response relationship (i.e. learning) which, in an electrical model, would probably require to be imitated by some autoselective switching device rather than regarded as time-constants of a resistance-capacity system. Another type of control demands the establishing of complex response-patterns which are 'triggered off' as a whole by the stimulus. Instances are the action of word-habits in typewriting, or of blocks of stimuli in transmitting morse, or of associated movement groups in knitting. These seem to require some 'sequencing' switchgear, of the type used in the Relay Automatic telephone system, and make us think of the physiologists' 'chain reflexes' and of rhythmic reflexes such as walking and breathing.

(*Manuscript received* 4 *March* 1945)