NOVEMBER 07

THE UNDERGROUND ELECTRIC TRAIN

by Piers Connor

29 – FROM SIGNALLING TO ATO

WHY ATO?

From the earliest days of railway operations, it was realised that individual drivers had individual driving techniques. The legions of stories written on steam locomotive performance are littered with descriptions of driver's handling techniques, their knowledge and understanding of their machines, the abilities of them and their firemen and critical analyses of their methods and results. Some drivers drove hard and fast while others were more cautious. Some drove on the regulator, some on the cut-off. Some drove so hard they pulled the fire from the grate and blasted it out of the chimney while the fireman struggled to keep up. Others consistently failed to keep time but gave their fireman an easier job. Whilst diesel and electric traction doesn't produce quite the same variety nor create the same opportunity for legends there are still wide varieties in driving techniques. During my years on the trains, we always had a few drivers who were known for their speed and others known for their lack of it, for their skill in braking or their lack of it and even guards who were slow or fast on the doors. My favourite nickname was the guard at Northfields who was dubbed "Speedy Gonzales"[1]. His train was always being held for time. You can imagine why.

Whilst these variations in technique were regarded amongst the crews as simply personal quirks, local problems of operation or minor inconveniences to timekeeping, the real problem, on densely trafficked urban routes in particular, is that variability in train performance leads to faster trains getting bunched behind slower ones while gaps in the service appeared ahead. A train with a long gap in front of it will soon become overcrowded and lose time as it struggles with the extra passengers gathering at each station. Although most crews know about these difficulties and accept that it is a problem, they believe there is little they can do about it on their own.

Other staff are involved too. Signal operators are supposed to be aware of such problems and are supposed to try to sort them out. The simplest response is to "regulate" the service by holding trains to time or by holding a train which is on time to make it late to help reduce the gap behind it. For this the signal operator has to pay attention, something which they are not all very good at. It requires a lot of concentration for a long time, an encyclopaedic knowledge of the timetable and an understanding of the continuous need for maintaining an even interval service. It would be wonderful if we could say that all signal operators have these qualities but, as with train drivers, it's not a perfect world.

A competent railway manager would not consider it acceptable for his service to be allowed to deteriorate because of the personal quirks of the staff. As a result,

---

1 I can't remember his real name, but his nickname was after the title of a popular song of the early 1960s by Pat Boone.

various attempts to overcome irregularities by the use of automatic devices of one sort or another were, and still are tried from time to time. In the early 1920s, an attempt to improve train interval regulation appeared in the form of a headway indicator adjacent to the District tracks at Earl's Court which showed the time a train was late or not as appropriate. This was manually operated by the signalman and was supposed to be read by the driver and acted upon appropriately. Its lack of success may be deduced from the lack of its adoption by other lines.

A further development was the "headway clock" introduced at a few stations on the Piccadilly Line. This showed automatically the driver the time in minutes since the previous train left the station. It assumed that drivers knew what the headway should be and that they would take the appropriate action to keep to it. Of course, few did, in either case. Headway clocks were still in operation when I worked on the Piccadilly in the mid 1960s but most of the crews ignored them.

Eventually, programme machines were introduced to replace most of the signalmen controlling and regulating trains. Although they were successful in reducing the number of staff required to operate the railway and they allowed some consistency in timekeeping, they could not remove the effects of variable driving techniques.

All these approaches relied on the crew or the service regulator to play their part. The faith is rarely justified by the results. The best solution is to take the chance factor of human response, or the lack of it, out of the loop – in other words, to automate. Increasingly, railways have adopted automation in power supply, traction control, train routing, maintenance planning and train operations. This automation of railway systems started with signalling and the Underground was one of the pioneers of it. Over a period of 60 years, they developed it, expanded it and eventually became the first railway to design and build a complete line with Automatic Train Operation (ATO).

On and Off

Amongst railway people, signals are said to be "on" or "off". "On" means the signal is showing a stop indication while "off" means it is showing a proceed indication. These descriptions refer back to the very early days of railways when a signal consisted of a board on a post. If you wanted to tell the driver to wait, you placed the board on the post, facing the driver, so the signal was "on" the post. When it was OK for the driver to go, you took the board "off".

Even today, some main line drivers still refer to signals as "boards". In some places, like the Midlands, they are called "pegs". On the Underground

AUTOMATIC SIGNALS

Automation of signalling came to the Underground early. An automatic signal, operated by the passage of trains instead of by a signalman using a lever, was tried on the Hounslow branch by the District Railway in 1901. It was imported from the US (of course) and it was successful enough for a complete system of them to be tried on the Ealing & South Harrow Railway (the Rayners Lane branch as far as South Harrow to you and I) when it was electrified in 1903. This too was successful and the Underground adopted it for its new tube lines and for the electrification of the District Railway. The Metropolitan also adopted automatic signalling eventually but it was never fully integrated with the other lines until taken over by London Transport in 1934.

BLOCKS

Signalling was introduced as a means of preventing train collisions. The basis was that you needed to give the driver of a train early warning of another train stopped ahead of him because of the low braking rates that can be achieved with the steel wheel/steel rail interface. You may recall how, back in Article 6[2], I described the problems with braking and adhesion and I mentioned that 8% adhesion was the standard used by the main line railways for calculating an emergency stop. This is pretty low (compared with at least 80% on a road vehicle) and we could reasonably expect to get at least 20% adhesion in the dry tunnels of the Underground. In the open, the current adhesion limit assumed for ATO on the Central Line is 6.5%. All this aside, what we know is that, since trains can't stop instantly or even very quickly, you need advance warning when you are closing in on a train ahead.

Early systems used to signal trains from station to station, the line between two stations effectively becoming a section[3] or, as it is described on main line railways, a "block". A train was only allowed to leave the station at the entrance of the block when it was confirmed that the train ahead was clear of the next station. A visual signal was placed at the entrance to the block to show the driver whether it was OK to proceed. Soon, a rule evolved, which is still valid today – "only one train is allowed in any one section at any one time". Obviously there are waivers for coupling, rescue and the like but that's the basic rule.

Gradually, this arrangement was developed so that the signalman controlled the area around his station and the admission of trains into the sections approaching his station. The whole operation was manual and relied on the vigilance of the staff to make it work safely. The signalman (nowadays called a "signaller", in case he's a she, or on LU, "Service Controllers") had to watch out until he saw the train depart his area of control before he would pass the "train out of section" message to the signalman in rear and subsequently "accept" another train when offered it.

Over many years of development, the signalman was provided with mechanical interlocks and "instruments" of various types which helped to prevent him from making some basic errors and to remind him of the status of the blocks he controlled. This worked OK as long as everyone was paying attention and things didn't get too busy but a number of accidents over the years showed that it wasn't foolproof. Signalmen could, and did, occasionally let trains run into occupied sections with sometimes fatal results. It was from these incidents that the idea of some sort of automatic train detection was thought desirable and, by the late 19th century, it had arrived in the form of the track circuit, the principle of which is shown in Fig. 1 below[4]. The system automatically shows that a section is free by completing a circuit. A train is detected when it occupies the track and shorts out the circuit. The system is used to operate the signal protecting the section. Originally it was battery operated, using direct current but it later evolved into a more sophisticated system using AC.

[2] *Underground News* No.528, December 2005.
[3] On a standard two-track railway, it is two sections, one for each direction.
[4] There were other systems, like treadles operated by wheel flanges but the track circuit was the most effective.

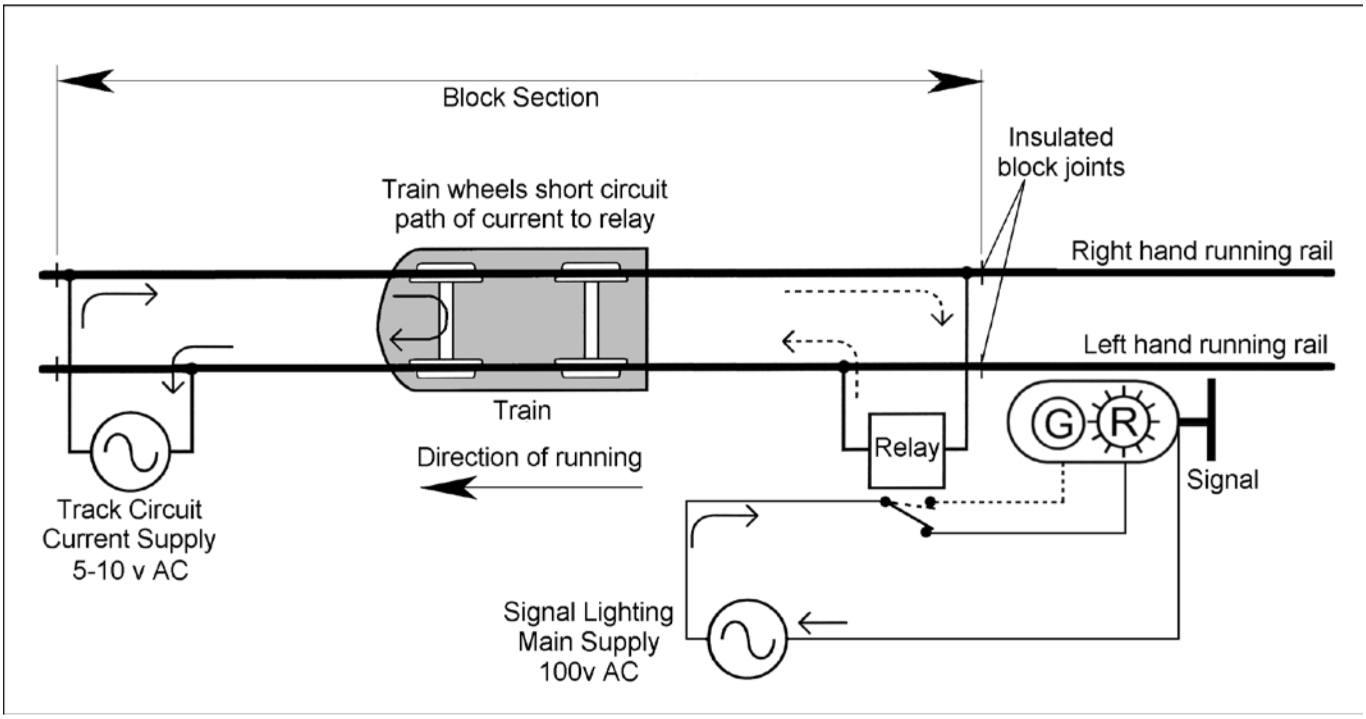

Fig. 1: Schematic of simple track circuit showing the arrangement for automatically operating the signal protecting the section. The diagram shows the section occupied by a train and the signal at danger. If there was no train in the section, the circuits would operate as shown by the dotted lines and the signal would show a green aspect. In reality, many sections only have an insulated block joint in one rail but the principle is the same.

If you pass a low voltage (5-10v AC) electric circuit through the running rails you can use the circuit to operate a relay to indicate current is flowing. If current is flowing the relay can be used to switch on a light – in our case a green signal lamp lit by 100v AC. If the current stops flowing, the relay contacts drop to switch the feed to the red light. The circuit is normally arranged so that the feed is at the far end and the relay at the entry end of the circuit. This gives continuity to the detection ahead of the train while it passes through the section.

Since the wheels and axles of train are steel, they short circuit the track as soon as they enter the section and you can use the loss of the circuit to show the section is occupied. The track circuit "goes down" as they say, the relay drops open and the signal shows a red light to the driver of the next train to approach. When the train leaves the section, the circuit "picks up" and the relay is energised and switches the signal to green. If you divide the line up into sections and isolate each section electrically from its neighbours by means of an "insulated block joint", you can protect the whole line with this system.

The arrangement was adopted by the Underground in 1903 and, eventually, by many other railways and it is still the principle form of train detection today. Modern versions have become available, for example where adjacent circuits use different, electronically generated frequencies and avoid the need for insulated joints – jointless track circuits or JTCs.

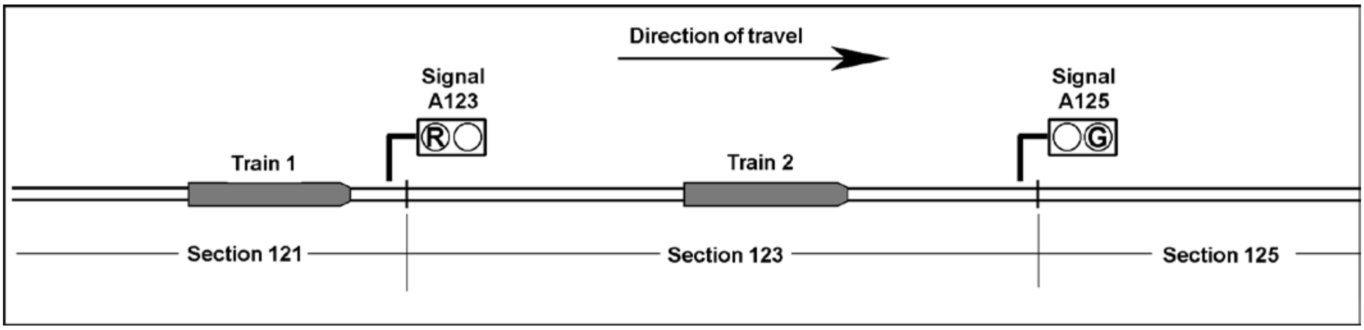

Fig 2: Schematic showing basic automatic block signalling adopted by London Underground. A signal is provided to protect the entrance of each section. Note that automatic signals on the Underground are lettered "A" and numbered like houses in a street, odd numbers one side, even numbers the other.

An additional feature of the Underground's signalling system, installed at the same time as automatic signals, is the trainstop. This, as we have seen before in this series[5], operates in conjunction with the signal, being raised to stop a train which attempts to pass if the signal shows a stop aspect. The train is stopped by its tripcock causing an "irretrievable" emergency brake application, i.e. the tripcock cannot be reset until the train has stopped. The use of automatic signals with trainstops for over 100 years has made the Underground one of the safest railways in the world.

OVERLAPS

Now that each section is protected by a stop signal and a trainstop, you might be forgiven for thinking that trains were now safe from being run into from behind. Unfortunately, this is not so. If a train passes a signal at danger and gets tripped, it needs lots of room to brake before it will come to a stand. If the train in the section protected by that signal is just inside the section, a short distance beyond the signal, the tripped train will hit it before it stops (Fig. 3 below).

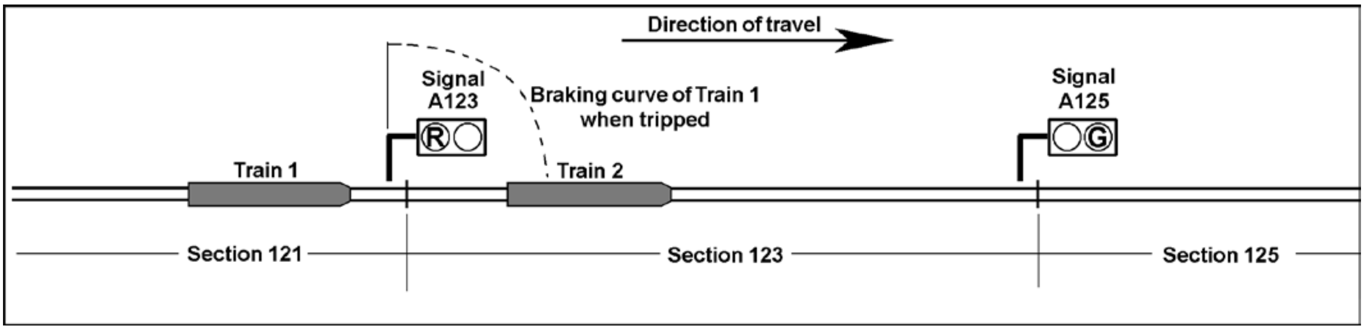

Fig. 3: Schematic demonstration of how a tripped train could enter an occupied section and collide with a train ahead. This is prevented by what is called an "overlap", Fig. 4 below.

To remove this possibility, you have to give the tripped train room to stop before it enters the section. This means it must be tripped before it reaches the section; in fact, it must be tripped a full braking distance before it reaches the section. So, to provide this distance, you have to move the signal and trainstop a safe braking distance back from the entrance to the section. The distance the signal has to be moved back is called the overlap (Fig. 4 opposite, top).

5 Article No.17, *Underground News* No.539, November 2006.

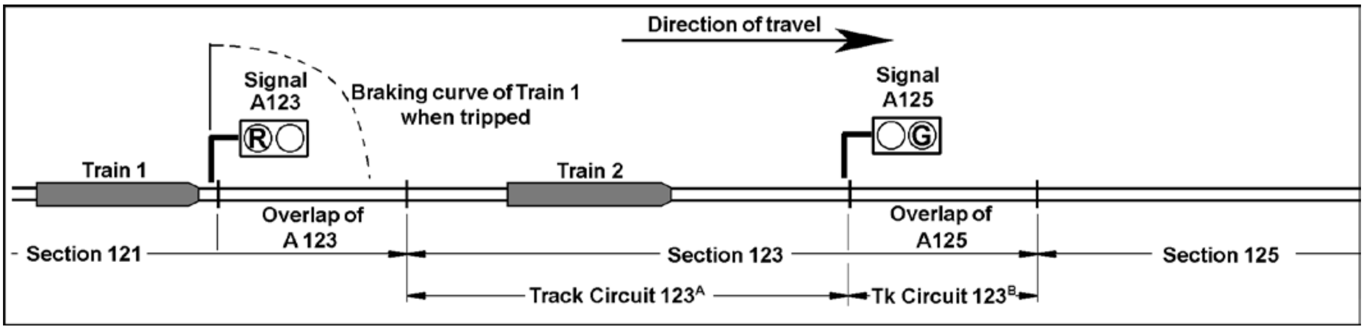

Fig. 4: Schematic showing how signals are positioned a safe braking distance back from the entrance to the section to allow room for a tripped train to stop. This distance is called the overlap. Note that the overlap has its own track circuit to ensure that the signal returns to danger as soon as the front of the train passes it.

Overlaps are calculated on a site by site basis. Over the years, a complex formula has evolved which takes into account the gradient, the maximum possible train speed, the braking capacity, a margin for error and even the position of the tripcock relative to the front of the train.

In some locations, the overlap is longer than a train's length, so it would be possible for the train to pass the signal entirely before entering the section that the signal is protecting. Until the train enters the section and the track circuit "goes down", the signal shows a green aspect behind the train, not something we want to do. To eliminate this possibility, the overlap has its own track circuit so that the signal will return to danger as soon as the leading wheelset of the train passes it.

This feature is reflected in the standard Underground track circuit numbering system for automatically signalled areas so that, in our example of Section 123 above, the track circuits would be numbered 123^A and 123^B, the latter being the overlap of Signal A125. Both track circuits still form the section 123 and both must be clear for A123 to show a green aspect.

REPEATERS

As trains need a lot of distance to stop, drivers need to know well in advance about the condition of the signals ahead of them. In many cases they can't see the signal early enough to be able to stop at it if it is showing a danger aspect. This problem was recognised very early on in the development of railways, so drivers were provided with "Distant" signals. The Distant[6] repeated the indication of the stop signal and gave the driver a good chance of stopping if he needed to.

The Underground adopted the same idea, but modified to suit its close headways and tunnel conditions. On main line railways using semaphore signals, distants are normally used in specific circumstances regardless of sighting but, on the Underground, repeaters are only provided where sighting of the stop signal is compromised in some way. Since the standard Underground stop signal is a two-aspect, red or green signal, repeaters for automatic signals show green or yellow[7]

6 When BR was trying to rewrite the Rule Book in simple English, the team engaged on the project hired in a "plain English" consultancy. The earnest folk from this agency spent some time trying to convince the railway people that a "distant" signal should be called "a signal which was far away". Such is the price one pays for allowing amateurs to meddle in the railway business.

7 There is no such colour as "amber" on the railway. It is always "yellow". From my perspective as a resident dinosaur, any servant of a railway company who refers to "amber" signals should be

aspects, depending on the aspect of the stop signal. Repeaters for automatic signals are lettered "R" plus the number of the associated stop signal.

Although a repeater is designed to repeat the aspects of its associated stop signal, there is a period when it doesn't. If a signal and its repeater are both showing green[8] and a train approaches, it will pass the repeater first. As it carries on, it is likely to be fully past the repeater before it reaches the stop signal. This means the repeater is showing green behind the train. To prevent this, like the overlap, the repeater has its own track circuit so that, as the train passes it, it changes from green to yellow. As a result, in the short time the train is between the repeater and the signal it repeats, the signal will show green while its repeater shows yellow, as shown in Fig. 5 below. So, you could say that a repeater repeats the aspects of its associated stop signal but, er, not all the time.

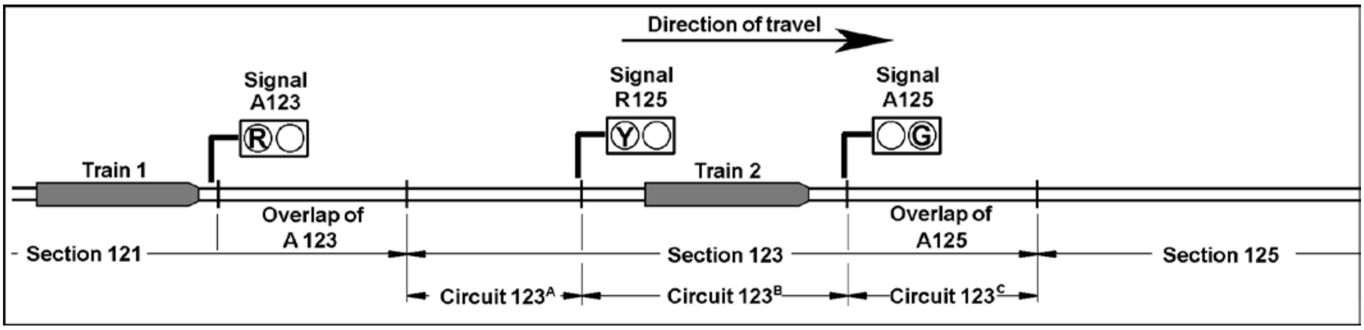

Fig. 5: Schematic of repeater signal showing how the repeater R125 shows a yellow aspect while its associated stop signal shows green using an additional track circuit. Section 123 now has three track circuits.

"Well", you might ask, "Why go to this trouble to prevent a green light showing behind a train? After all, there is a red signal protecting the section it occupies (A123 in Fig. 5), so there won't be any way the driver of a following train can see the green of the repeater." Well, actually, there is. It's called the "Stop and Proceed Rule".

STOP & PROCEED

It was discovered very early on that automatic signals don't always behave themselves. Now and then, they would fail to clear after the train had left the section they were protecting. The "signal failure", bane of the commuter, had arrived. Actually, a signal staying at red when the train isn't in the section is not actually a failure. It is the system reverting to the next level of safety. Something hasn't done what it was it is supposed to do so the signal isn't about to let trains go past it to enter a section which might be occupied. It stays at red. Inconvenient but safe.

In the early days of Underground electric train operation, there was no radio and few phones. Trains in tunnels were effectively isolated from the rest of the world. To allow some sort of contact to be made in the event of a signal remaining at danger, the stop and proceed rule was introduced. This rule told the driver to wait one minute (nowadays two) and then proceed past the signal slowly, basically driving on sight. Of course, the train would be tripped. The driver had to reset the tripcock and then move off slowly into what might be an occupied section.

---

signed off as medically unfit and should only be allowed to return to duty when he can certify, in writing, that his brain has been thoroughly washed!

8 A repeater will show green when the signal it repeats shows green and its trainstop is lowered.

This process achieved two objectives. It allowed a train stuck at a red signal to move forward to the next station at least and let everyone know at "Euston, we have a problem". Also, it provided a means for a stalled train to be approached and perhaps, be assisted by the next train – the classic "push out". Both have been used, the former the most regularly. Today, with train radio, the process is a bit easier. At least the driver can call up the control room and ask if they know what's going on.

As we've seen before[9], the driving on sight after "tripping past" a suspect red signal has not always been too successful. There were collisions, sometimes fatal, so "Speed Control After Tripping" (SCAT) was introduced. This enforces a 10mph speed limit on the train for 3 minutes after it has been tripped.

For our situation where a train is in the section ahead and its full length has passed a repeater, a following train "applying the rule" should not be shown a green aspect. This is further backed up by additional diktat which prohibits resumption of full speed until two STOP signals have been passed showing a proceed aspect.

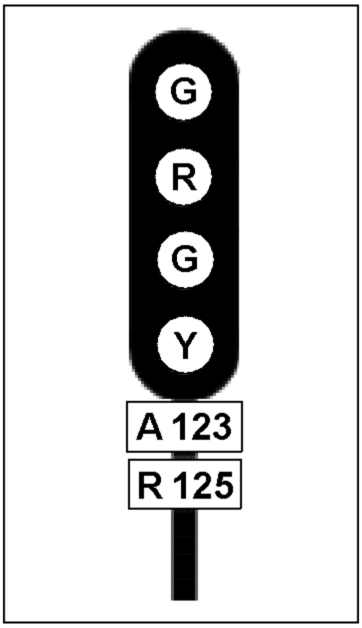

Fig. 6: LU signal post carrying a stop signal above the repeater for the next signal. Note the two identification plates.

FOUR ASPECTS?

Because you can only have one train in one section at one time, if you want to let more trains through the system, you have to reduce the length of the sections. This will mean that successive signals will be closer together. Sometimes, this can mean that the signal for the section you are approaching has the repeater for the next stop signal on the same post. Semaphore signals are also seen like this – the distant for the next signal is mounted below the previous stop signal.

On the Underground, particularly in open sections, the repeater and previous stop signal were often be made up as a single signal unit with four aspects. If all the signals were clear, the four aspect signal would show two greens – green for the stop signal (always at the top) and green for the repeater of the next signal. The signal head is as shown in Fig. 6, right.

This design should not be confused with the 4-aspect signals used on the main line. The main line 4-aspect signal indicates the state of the road over four sections. The sequence of operation is red, single yellow, double yellow and green. Also, the main line 4-aspect signals have only one identification plate. The Underground version, actually being two signals, has two number plates, one in white for the stop signal above one in yellow for the repeater.

Certain signals on the Metropolitan Line north of Harrow-on-the-Hill are genuine 4-aspect signals, installed to provide sufficient warning to main line trains operating with longer braking distances than usual on the Underground. These areas do not have repeater signals, since they could be confused with a multi-aspect yellow. If a repeater is necessary because of sighting problems, a banner signal is used. Rickmansworth southbound comes to mind as one place which has a banner signal.

Incidentally the Underground, where signal design policy long ago forsook aesthetics for economy, has abandoned the 4-aspect look-alike stop signal and repeater

9 Article 17, *Underground News* No.539, November 2006.

combination as a single signal head and now they consist of two separate 2-aspect signal heads, even if they are on the same post.

TRAINSTOP PROVING

A feature of London Underground signalling is double protection. Nothing is ever allowed to be exposed to a single fault which could render it unsafe. Trainstop operation is a good example. A trainstop is the ultimate protection for the train. When a signal shows a stop aspect, the trainstop is raised to trip into action the emergency brakes of a train which passes the signal. The trainstop is lowered only when the signal shows a proceed aspect.

Now, the trainstop is raised by a spring and held down by compressed air pressure. This means that, if the air pressure is lost, even if the signal is clear, the trainstop will automatically rise. The signal control circuit detects this and causes the red aspect to appear, even though the section ahead is clear and the green aspect is lit. The driver will then see both red and green together - the “dual aspect” – which he will treat as a stop signal. He then applies the “stop & proceed” rule.

The other possibility for a trainstop malfunction is that it doesn’t go up when the signal is red. It has happened – suppose an empty can gets wedged between the trainstop and its control box – so a protection circuit called “trainstop proving” is used. This circuit prevents the signal in rear of our signal from clearing if our signal is red but the trainstop hasn’t gone up. This is known as “trainstop proving”. This kind of protection is used frequently in signalling circuitry and is nowadays called “diversity” – never have a single point of failure, always check everything twice.

IN REAR AND IN ADVANCE

When discussing operations on a railway, particularly when you are dealing with an emergency, you have to be sure you know where everyone and everything is. Over the years, expressions have developed to indicate locations of trains, signals, structures and people in relation to other objects along the line. Sometimes, to the outsider, these expressions can be confusing. Two such are “in rear of” and “in advance of”. The following diagram, Fig. 7 below, shows what these two expressions mean.

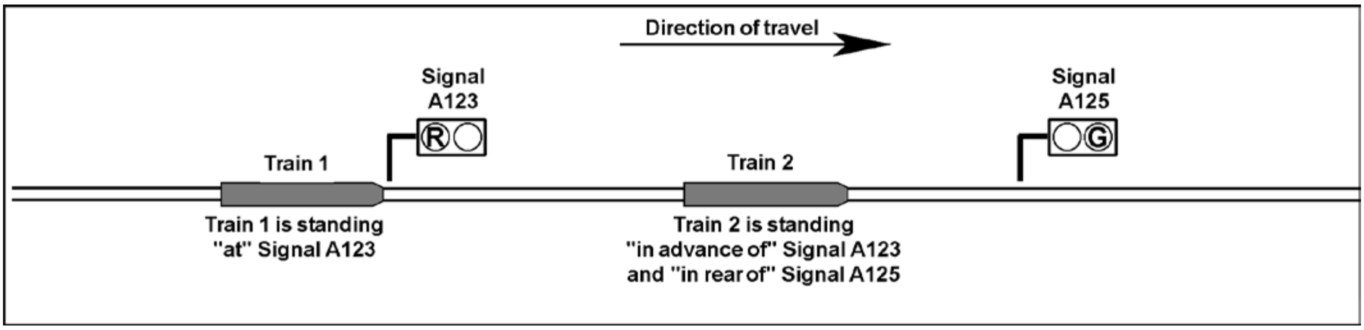

Fig. 7: Schematic of section of line showing descriptions of locations.

Train 1 is standing at signal A123. This is how the driver would report his position when contacting the signaller. This is simple and could hardly be misunderstood. However, Train 2, stopped between signals A123 and A125, has passed signal A123 but it has not reached A125. Although A125 is ahead of the driver as he sees it, he is said to be “in rear of” the signal. And, although it has gone past signal A123, the train is said to be “in advance of” signal A123.

If you aren’t paying attention, this could get confusing. It is made even more confusing by the appearance in LU Signal Engineering standards of a statement which says that the stopping point for a train “shall normally be 5 metres IN FRONT OF a signal” (my capitals). If this was true, i.e. in advance of the signal, the train would have been tripped! Of course, the signal engineer is looking at the signal the other way round. The back of the signal is where he changes the bulbs and feeds the wires in. The front is where he cleans the lens.

This has the hint of farce about it but there have been recent instances where errors have been made because of the misinterpretation of these expressions. With the huge number of new people coming into the railway sphere these days, it is essential that these descriptions are standardised, communicated during training and verified as clearly understood when used on the ground. Unfortunately, it isn’t always done.

To be continued ……