Nuclear weapons, AI, and the June 1980 US false alarms

Artificial Intelligence (AI) and its potentially dangerous link to nuclear weapons is a quite popular topic these days. Some of it is because nobody really knows what the role of AI can be. It is a (relatively) new and quite exciting technology and it seems to offer some benefits in assessing complex situations. People also intuitively understand that there is a potential danger there - since we don't (normally) know how an AI system arrives to its conclusions, we cannot be certain that its assessment of a situation or its recommendations are correct (whatever that means).

Given that when is comes to nuclear weapons the cost of an error can be rather high (which is a serious understatement), one can easily see why people usually like the idea that, quoting Antony Blinken, "artificial intelligence should not be in the loop or making the decisions about how and when a nuclear weapon is used".

This is all good, but the issue seems to be a bit more complex than just keeping a human in the loop (or AI out, for that matter). After all, people make their decisions regarding nuclear weapons based on information that is processed by computers one way or another. Whether some of this processing was done by AI is not necessarily important, in my view. All technologies have some magic element to them, and it's possible that in the case of AI, the difference is just a matter of degree. This difference may be important or it may not be. I must say I haven't done a literature review, so if there is a good analysis of the role of AI in nuclear decisions, please send me the link.

I believe that one of the issues we are dealing with here is the interaction between people, organizations, and information. And we have some examples of how this interaction works, specifically when it was about real-world decisions regarding nuclear weapons. Not many, but some. And they can tell us something. One of these examples is the series of US false warning accidents in June 1980. This is not the only one accident, of course, and it has been studied by others in some detail. (Again, if you have a particularly good study in mind, please send me the link. Scott Sagan, of course, covered it in "Limits of Safety," but he was looking at the safety aspects of the event and he had no access to the documents declassified by the National Security Archive since then.) This is an attempt to reconstruct the June 1980 accidents, largely to get a better picture of what happened.

Documents

The National Security Archive has done a great job compiling a set of declassified documents in its collection "False Warnings of Soviet Missile Attacks Put U.S. Forces on Alert in 1979-1980." There is also the Hart and Goldwater report to the Senate (HGR below). There are some differences between different accounts, some of which will be noted below.

The key documents from the NSA collection used here are the Secretary of Defense Harold Brown's memo to President Carter from 7 June 1980 (HB below), the talking points prepared by the Office of NORAD Deputy Chief of Staff dated 21 July 1980 (NRD below), and the Department of Defense "Fact Sheet" (DoD below)

Background

This largely follows the Hart and Goldwater report that provides a useful summary of the situation and the key players.

There were four key command centers, the main one being North American Aerospace Defense Command (NORAD). Then there were two National Military Command Centers - the National Military Command Center (NMCC) in the Pentagon and the Alternative National Military Command Center (ANMCC), located in Fort Ritchie, Maryland. There was the command center of the Strategic Air Command, at the Offutt AFB, Nebraska.
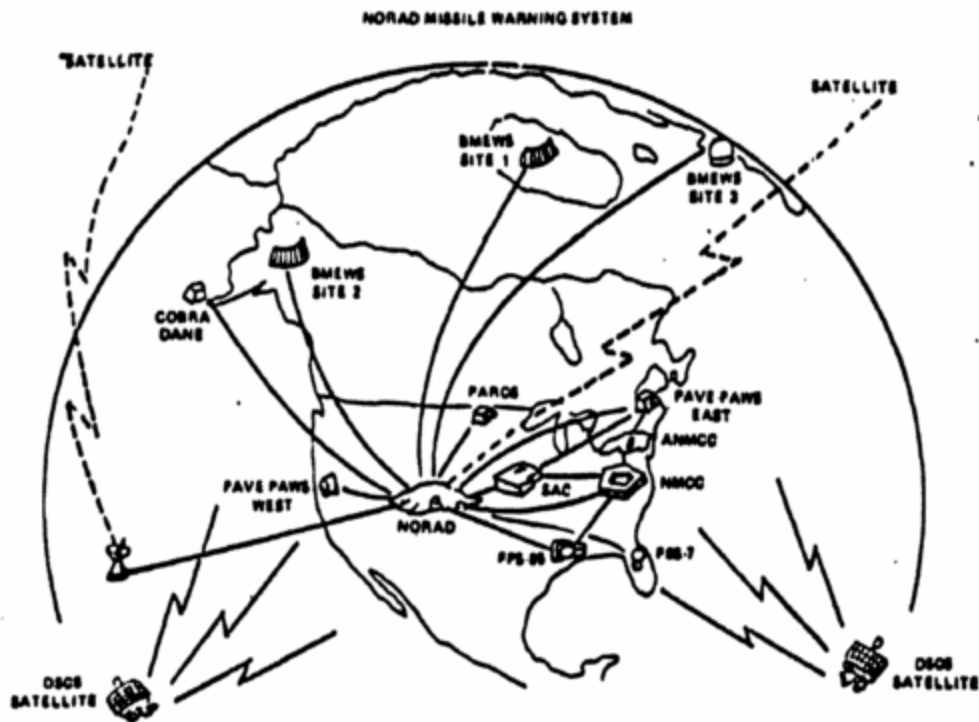
This schematic representation of the structure of command centers and sensors is from HGR. It does not seem to be entirely accurate, but it gives you an idea. NORAD is receiving data from all early warning sensors - satellites and radars. The other centers are getting data directly from only those sensors that are tasked with detecting SLBMs - satellites and PAVE PAWS radars. NORAD provides the lower-level centers with its analysis of the attack data, presumably based on the entirety of information that it receives. This means that duty officers at the lower-level centers are looking at two sets of displays - one with direct data from some of the sensors and the other one with the NORAD analysis of the data from all sensors.

If there is an indication of a threat, the four centers initiate a conference call. There are three types of conferences:

The first one, Missile Display Conference (MDC), is rather common event. HGR reports that there were about 1500-2000 events a year that resulted in MDC in 1979-1980. So, these are very common events, and only a fraction of them, less than 100 a year, is evaluated as potentially threatening North America.

One interesting detail (HGR) is that all 3000+ MDCs convened in 1979-1980 resulted from some external physical signal. But in addition to this, there were events where "a computer or a piece of communication equipment will transmit a false piece of information." These would not trigger an MDC and, indeed, were not recorded at all before June 1980. NORAD told HGR that these kinds of events happened about "two or three times a year." Who exactly determined that these are "random failures within the computer and communication equipment" and how this determination was made is not entirely clear.

It appears that any command center can start a conference, but it is not quite clear at what point others are asked to join. I would assume that once one center initiates an MDC, all centers, including NORAD, would be on the call. But there will be a question about this later.

The next level is a Threat Assessment Conference (TAC) that involves higher-level people (up to the Chairman of the Joint Chiefs of Staff). It appears that the TAC protocol involved "preliminary steps to enhance force survivability," such as launch of airborne command posts. The top-level conference is a Missile Attack Conference, with the President brought in. This has never been convened before 1980 (and probably after).

In addition, the Strategic Air Command (SAC) has its own procedure that runs in parallel to the conferences process. Even if the data are ambiguous, SAC can implement certain survivability enhancing measures. These measures include directing bomber and tanker crews to board their aircraft and start engines in preparation for take-off, if necessary. As can be understood from the accounts (and from the actions taken afterwards), SAC duty officer ("SAC duty controller") can make this decision on his own discretion, without waiting for any conferences to be convened.

28 May 1980

As it was determined later, the actual start of the chain of events was 28 May 1980 - displays at two command centers showed that a large number of missiles are coming - 2020 at SAC and 9000 at ANMCC. The signals appeared close together [HB], but "at different times" [NRD]. These showed up for only six seconds at SAC and for similarly brief period at ANMCC. No conference was convened by any of the centers. Apparently, these were dismissed as one of those rare "random failures." Not entirely dismissed, though - NORAD

began a technical investigation of anomaly. It may well be that this kind of investigation follows each of the 2-3 per year "random failures," but there is no indication of this in the documents.

3 June 1980

The first real accident happened on 3 June 1980. Reconstructing the timeline is a bit difficult. HGR says the first indication appeared "at approximately 2:26am Eastern Daylight Time." NRD places the first "threat" display at 0525. I'm not sure what time zone it is as it doesn't look like GMT or UTC. And definitely not the time in Colorado, where the NORAD HQ was located. It's not particularly important, though and I will use the EDT time zone from HGR but the minutes from NRD. After all, it produced the famous "3am phone call" meme.

In fact, according to NRD, there was an event a bit earlier - at 02:21 NMCC registered "non-threat missiles." Since these were "non-threat" and "the display cleared within seconds" no action was taken. From this one can conclude that there is a separate display for "non-threat" missiles but this point is not elaborated in the documents.

02:25 The accident proper began at 02:25, when the SAC display showed 2 "threat missiles" identified as SLBMs. Eighteen seconds later the SAC display started showing 200 SLBMs [HGR]. According to HGR, at this point SAC personnel called NORAD regarding this display and NORAD told them that they had no indication of any launched SLBMs. However, no other account has this call. One would guess that a call like this would be made as part of a standard Missile Display Conference, but the conference will not be convened until later.

02:27 Whether or not the SAC-NORAD call took place, SAC made a decision to act and "about two minutes later" [HB] or at 02:27 [NRD] "SAC duty controller" directed "all alert crews to move to the alert aircraft [bombers and tankers] and start their engines" [HGR]. As can be understood from the NRD account, sending crews to their aircraft involved what was known as "Fast Klaxon Alert."

02:28 According to NRD account, NORAD issued "ALL CLEAR" - no detection by any sensors. Again, it is not clear how exactly this determination was made and how it was communicated and to whom. It appears that it was still a phone call between NORAD and SAC. HGR also says that at some point "NORAD showed all systems clear."

02:29 SAC issued a directive to assume "Posture 5" [NRD]. The HGR account says that shortly after the order to start engines was issued, the SAC display cleared and "shortly thereafter" the SAC crews were directed to shut their engines but remain in aircraft. This may have been the "Posture 5" order. Note that HB mentions none of this - neither the clearing of displays nor the directive to shut the engines. The bottom line is that the SAC displays cleared and the aircraft engines were shut. DOD states that the decision to shut down aircraft engines was made by the "Commander in Chief, SAC."

02:31 According to HB, six minutes after the first display reading, SAC "began displaying 2020 ICBMs from BMEWS." Note that these were not SLBMs, so SAC would not have a direct signal from sensors, just a display that was supposed to show the information coming from NORAD.

02:37 Twelve minutes after the start of the accident, the display at NMCC began showing 200 SLBMs coming to the United States [HB].

02:39 At this point [NRD], NMCC Duty Officer convenes a Missile Display Conference [HGR]. It's not clear what happened to the SAC display that showed 2020 incoming ICBMs at 02:31. HB reports that "shortly thereafter" (after MDC), the display, probably the one at NMCC, started showing 220 ICBMs. That would be different from the 200 SLBMs displayed earlier. What's interesting is that, according to HB, the display showed that all 220 ICBMs "had already impacted."

02:39-02:49 These must have been some really tense ten minutes. It's not clear what happened to the displays but they probably continued to show something. According to HGR, "all command posts were convinced that the data were erroneous and invalid."

02:49 The NMCC Duty Officer upgraded to conference to a Threat Assessment Conference (the time is from NRD). This decision is hard to explain. If everyone was convinced that the data were erroneous, why upgrade the conference? HGR says that it was done "as a way of terminating the incident and insuring that all parties knew that there were no threatening activities." This is way too generous, in my view. None of the reports asks why the decision to terminate the incident could not have been taken at much earlier time, even before the MDC, and why it required going all the way up to a Threat Assessment Conference. HB says that at the time TAC was convened, "a NORAD assessment was requested." But according to NRD, "NORAD assessed information as FALSE" at 02:40, shortly after the MDC was convened.

02:51 A Threat Assessment Conference is a serious business. It triggers all kinds of actions, one of them is a launch of Airborne National Command Post (ABNCP) aircraft. According to NRD, the command post of the Pacific Command took off at 02:51. The National Emergency Airborne Command Post (NEACP) aircraft at Andrews AFB, was also preparing to take off. According to DOD, the NEACP aircraft "initiated prescribed actions preparatory to emergency launch" when it received notification of the threat of a "large number of SLBMs." That would be the very beginning of the incident, 02:25, when 200 SLBMs were "detected." DOD further reports that the NEACP aircraft "taxied in position for emergency launch," but by that time the SLBM treat was no longer there, which appears to be the 02:29 point, when the SLBM display cleared. At that point, "senior watch officer directed NEACP to hold in place" [DOD]. William Odom in his note to Zbigniew Brzezinski said that "NEACP didn't move," but

he probably meant that it didn't take off. It is reasonable to assume that both aircraft received the same order at the same time, but one took off while the other didn't. It appears that the Pacific ABNCP had a different "senior watch officer."

02:54 NORAD issues a formal "NO" output [NRD]. This seems like something everybody was waiting to happen, but it's not clear how this "NO" was different from those that NORAD repeatedly communicated earlier - see the SAC-NORAD call at 02:25, "ALL CLEAR" at 02:28, and the assessment of information as FALSE at 02:40. (In parenthesis, the NRD account also mentions that "CINCLANT incorrectly reported SLBMs indications" at 02:53, but it's hard to see how this fits in the narrative as these indications did not originate from any of the command posts or sensors.)

02:57 Finally, the Threat Assessment Conference is terminated, 32 minutes after the first erroneous reading appeared on the SAC display.

02:58 SAC, of course, had a separate procedure and took another minute to terminate its own alert.

So, the accident was over 33 minutes after it began.

6 June 1980

In response to the 3 June incident, NORAD took some measures, such as communicating with all users, analyzing the data, and installing some "software trap" (whatever that means). At the same time, it also placed computers back to the 3 June configuration [NRD] and tried to "duplicate the error ... hoping to reproduce the erroneous data" [HGR]. Whether as a result of these efforts or by happenstance, the error did appear again.

The 6 June timeline is more confusing. HGR says that the first erroneous reading was displayed at 3:38pm, probably Eastern Daylight Time. The NRD account places the first display at 19:49. Not only the time zone is different, but the minutes are different as well. Since the absolute time doesn't matter that much, the description below will follow the NRD times.

19:49 As reported by NRD, the display at SAC showed 2000 incoming ICBMs. HB adds that this was followed "in a few seconds" by displays of 6 missiles at NMCC and 2020 missiles at ANMCC. NRD also mentions the NMCC and ANMCC displays, but doesn't say anything about them other than that they were there. HGR reports that these were ICBMs as well. HGR doesn't say it directly, but one gets an impression that operators, at least at NORAD, were aware of the effort to reproduce the error.

19:51 At this time, SAC again sounded its Fast Klaxon Alert, which sent crews to their aircraft and directed them to start their engines [NRD, HGR]. HGR reports that they were directed to shut down engines (but remain in the aircraft) shortly thereafter.

19:51 While SAC does its own thing, NMCC convenes a Missile Display Conference. During the conference NORAD apparently says that there are no signs of an actual attack [NRD]

19:52:30 Here is the moment when NORAD formally says "NO" [NRD].

19:53 MDC is formally terminated. That was four minutes after the first displays at SAC, NMCC, and ANMCC.

20:03 That's when the "spurious indications" finally stopped [NRD].

20:06 According to HB, it took SAC 13 more minutes to terminate its alert - "17 minutes after the first false display appeared at SAC" [HB]. It appears that SAC waited until everything is truly over.

After the incident was over, NORAD switched its computers. According to DOD, the backup computer used "completely different hardware and software from the system that caused the problem." (In parenthesis, the May 1981 GAO report "NORAD's Missile Warning System: What Went Wrong?" says, apparently incorrectly, that there was a Threat Assessment Conference on June 6, 1980).

Some lessons

This part of the story is reasonably well known - the culprit was a 47 cents computer chip that randomly inserted number 2 into the communication signal that went from NORAD to lower-level command centers (although this doesn't quite explain the 6 missiles display at 19:49 on 6 June or 9000 missiles display on 28 May).

It's certainly worth noting that the idea of constantly sending a message that would show the number of incoming missiles was designed as a clever feature. That message supposed to "to have a continuous check on the condition of the circuits." It wasn't quite a safety feature, but it was designed to make the system more efficient in some way. But, of course, what it did was it introduced a new point of failure (NORAD changed the format of the message after the incident).

So, what does this accident tell us? Most importantly, in my view, is that having a human in the loop does not really guarantee anything. Human operators have their checklists to go through and procedures to follow. It is telling that even though operators had access to the data from satellites and radars that were telling them that there is no attack underway, they still put a lot of weight in that display that showed them some random numbers. Imagine that it were not a numeric display but some situation assessment screen generated by a presumably sophisticated AI algorithm. A call to the upper-level (NORAD) center may not resolve the situation if that center does not really know how that screen is generated.

The HGR authors asked what appears to be the right question, which is whether it is possible to develop a checklist that would cover all possible scenarios or to tell operators to find balance between reliance of checklist and application of sound judgement. They left the question hanging but seemed to be skeptical about a comprehensive checklist. And they had a point.

It appears that they had more faith in the balance between checklists and "sound judgement." But that's not quite straightforward. First, the entire episode showed that it is too easy for the checklist to prevail. I do not see a reasonable explanation to why the episodes were not terminated much earlier. If everything went as described, the Missile Display Conference should not have been convened on either June 3rd or June 6th. And there is no reasonable explanation for the Threat Assessment Conference on the 3rd.

It may have helped had the checklist included the option like "I don't trust this thing and have no idea why it behaves like this." But it didn't and the operators apparently did not feel that they create one on the spot.

The SAC procedures deserve a separate mention. HGR says that SAC's decision to send crews to their aircraft "was not an untoward move." Perhaps. It was arguably a reversible step, which didn't have an escalatory potential on its own. But one would hope that someone would think about sounding that Fast Klaxon Alert on June 6, when everybody knew that the system generates erroneous messages. And definitely it's hard to justify waiting for 13 minutes after the June 6th MDC concluded that the alert is over.

The accident showed that the assumptions that go into your procedures matter a great deal. If your assumption is that the Soviet Union can attack you any minute you design your procedures accordingly. You can try, of course, to include certain safeguards, but in the end this assumption will dominate your thinking and your checklists.

And, more importantly, it will affect your presumably "sound judgement." For example, HGR noted that "there seemed to be an air of confusion" after the alarm was over on June 3rd. Of course there was. But HGR authors worried about this confusion for an interesting reason. For them, it pointed at a possibility of what we would probably call a cyber attack these days. The scenario would be that the adversary would introduce some "stray and erroneous data" into the system, which would confuse everyone to the point of undermining the effectiveness of the entire system. If this is something that the operators are trained to keep in mind, one can see how this can seriously affect their judgement.

I don't really have a good bottom line here. Maybe it's that the technology that is used in nuclear command and control, whether it's something old-fashioned or sophisticated AI, doesn't matter that much. The basic design of the command and control network and the underlying assumptions that you made are more important. Yes, it is vital to keep humans in

the loop, but we shouldn't overestimate their role. If your human operator cannot conclude that nobody starts a nuclear war with five missiles, as Colonel Petrov apparently did (even if it's a bit more complicated), then having them there won't really help you.

Anyway, the issue of AI and nuclear weapons is going to be around for some time. I hope that this attempt to describe the 1980 accidents will help inform the discussion of the issue. I will also try to put together a post on what we can learn from the Soviet accidents (that we know about). It can be very interesting in my view.

Note 20.11.2023: The post was updated to add references to the DOD document.

TrackBack

TrackBack URL for this entry: https://russianforces.org/cgi-bin/mt/mt-tb.cgi/1585