Fearon’s (1994) original model that introduced the notion of “audience costs” is both less than what supporters of audience cost theory (ACT) think and more than what critics imagine. It is less because it does not specify the mechanism that would generate escalating audience costs but only studies their coercive effect assuming that they exist. It is also less because it makes some assumptions that can be shown to be distorting – it ignores private communication and other instruments of coercion. It is also less because its most popular conclusion – that democracies are “more credible” in crises – actually comes from unmodelled assumptions that connect regime type to the assumption about audience costs, and is thus doubly removed from having firm theoretical foundations.

The model, however, is also more than what critics allege because it is not wedded to a particular substantive interpretation of its parameters, and as such the fact that it assumes escalating audience costs makes the conclusion about their coercive effect independent of the generating mechanism: if such a mechanism can be found, then the conclusions automatically follow. It is more because it spurred research to provide such a mechanism by connecting domestic politics to crisis behavior. It is also more because it provided the foundation for the study of reasons leaders might try to avoid public commitments, opting instead for vague threats, and attempt to finesse crisis outcomes. Even if all of these activities undermine the effectiveness of the audience costs mechanism as a coercive device, the fact that leaders try to avoid incurring such costs suggests a real concern that they might (which is consistent with the theory), and furthermore suggests that the theory’s conclusions might be valid in those cases when leaders are unable to avoid these costs. This has sparked research of the conditions that can prevent leaders from freely manipulating public opinion.

Trachtenberg (2012) makes many valid points but the article ultimately fails as critique of ACT for two reasons. First, all the issues he raises have been made before. The absence of references to that work is surprising in an article whose goal is to engage seriously the theory. Second, the historical evidence presented for the central claim – that opponents do not perceive public statements as hand-tying – is slim and often unpersuasive.

If one were to interpret this article as taking part in the modeling dialogue, then the cases could provide a corrective to some of the assumptions of the original theory. There is much value in this kind of dialogue between theory and data, and any viable specification of ACT must take into account several of the issues Trachtenberg (2012) discusses. If, however, one were to interpret the article as a broad indictment of scholars who work on the “purely abstract level” and who easily “get carried away with a particular idea” so that they “lose sight of all the reasons the particular effect... might not count for as much in the real world”, then its claims are unpersuasive. Critiquing the theory while ignoring its current status goes contrary to the modeling enterprise and does injustice to the original insight,
whose explicative and generative usefulness has been staggering, not just in the number of references, but in the development of new ideas in the context provided by that theory.\(^1\)

I first discuss the status of audience costs theory (ACT), both as specified originally and as modified subsequently. This exercise is useful insofar as it addresses other critiques of the theory as well.\(^2\) I then turn to some specific claims Trachtenberg makes and take issue with their logic, their historical accuracy, and the assessment of the evidence.

1 What Does Audience Costs Theory Say?

Let me begin by agreeing with the critics. Four our purposes, the original ACT can be usefully summarized as follows:

\[
\text{If (i) backing down in a crisis makes an actor suffer costs in addition to those arising from conceding the stakes, (ii) these costs increase as the crisis escalates, (iii) these costs can become so large that war becomes preferable to a concession, (iv) there is no other mechanism for coercing the opponent, and (v) attempting to coerce the opponent does not increase his costs of conceding, then escalation can commit an actor to fighting, and the resulting risk of war discourages bluffing, which makes escalation informative and gives it a coercive role.}
\]

Setting aside requirements specific to the game-theoretic model – e.g., “common knowledge of all parameters”, the concept of equilibrium, the ruling out of “strongly optimistic off-equilibrium-path inferences” (Fearon, 1994, 582–3) and some obvious other domain limitations – e.g., a world with two actors and no future interaction – we are left with a list of at least five premises, all of which are assumed to be true, and all of which might be necessary for the conclusions to hold. I say “might” instead of “are” because additional analysis is needed to ascertain the importance of these assumptions, not to mention their empirical scope. It is important to note that there is nothing in the original theory that connects the type of political system to crisis dynamics. \textit{The original theory is silent on whether democracies are more or less able to commit credibly during a crisis.} This is what Fearon (1994, 582) calls a “plausible working hypothesis.”

The first three assumptions constitute the core of ACT. They are critical but without microfoundations.\(^3\) Why would an actor suffer costs for escalating and backing down? Fearon (1994, 581) justified the assumption with an appeal to \textit{domestic} political opponents who would punish such behavior because they “deprecate that \textit{international} loss of credibility, face, or honor.” Following this cue, most attempts to rationalize audience costs have tended to conceptualize audience costs as punishments of leaders who “fail to fulfill their commitments” (e.g., by failing to carry out a threat). Before exploring this approach, I should note that it is not actually necessary to commit to it. One can also conceptualize audience costs as punishments of leaders who fail to pursue a policy the audiences want, with the preferences of the audiences molded in part by the leaders who attempt to generate support for their crisis stance by persuading the public of the necessity of a firm course.

---


\(^2\) E.g., Snyder and Borghard (2011); Downes and Sechser (2012).

\(^3\) Smith (1998); Schultz (1999); Slantchev (2006).
Thus, if during the crisis the public progressively becomes more belligerent, the leader will find it more difficult to back down: the political costs of doing so would not arise because he “said one thing but did another” but because of the perception that he acted contrary to public interest. As we shall see, this logic, which in formal terms would have the equivalent effect of the usual one even though it would have very different microfoundations, can be seen in action in some of the cases Trachtenberg discusses. This logic appears to be the only one I am aware of that is capable of generating audience costs that escalate with the crisis.

The problem for ACT was to specify a mechanism such that (i) audiences find it optimal to punish leaders who back down, and (ii) leaders pursue strategies which result in such punishments in equilibrium. This is not the place to rehash some of the assumptions that are needed to get all ducks in a row. Suffice it to say, it is quite difficult to get leaders, political opponents, and audiences to generate audience costs when they behave strategically and pursue their interests (Slantchev, 2006).

The link between regime type and audience costs might even be less straightforward: Slantchev (2006) argues that both autocracies and democracies have problems generating audience costs – the former because the opposition is too fearful to reveal problems with the leader’s policies, and the latter because the opposition is unconstrained and can say anything for electoral purposes. Whereas this puts into doubt the “working hypothesis” about regime type, the study also points out that the existence of a free and unbiased media can be quite important in generating audience costs, more important even than the political structures per se.4 Potter and Baum (2010, 466) argue that by ignoring the role of the media, the original ACT has “limited empirical support” and go on to observe that there is a trade-off between the leader’s ability to frame domestic political debate and their capacity to commit on international issues. Another strand in ACT also endogenizes audience costs but shows that it is quite possible for non-democracies to generate them: unpopular policies can destabilize these regimes, and their leaders tend to face more severe penalties when removed from office.5 Thus, the higher costs of losing can offset the lower probability of the sanction being applied, resulting in an overall expected risk that might be similar to those run by democratic leaders. Either way, it should be clear by now that the notion of “democratic credibility through audience costs”, while indeed plausible, turns out to be deeply problematic, and that it was recognized as such by scholars working out the theory itself.

I will not dwell on the third assumption except to point out that Fearon (1994, 584) is explicit about it, and that Slantchev (2011, 51, 61) both notes how important it is for the results, and doubts its empirical plausibility to the point of bluntness: “both [domestic audience costs and reputation] provide leaders with levers that they barely have access to in practice. As a consequence, the ability to commit derived from such devices is suspect.” The situation gets even murkier once other instruments of coercion are considered (excluded by the fourth assumption). For example, in the context of military threats, it is quite possible for crises between two high-audience-cost actors to have the highest probability of war (Slantchev, 2011, 236-8). The problem is that whereas it is true that high audience costs

5Weiss (2012); Weeks (2008); Bueno de Mesquita et al. (2003).
make escalation more dangerous and so less likely, the increase in the risk of war they create might offset the decrease in the probability of escalation, leading to an overall higher risk of war. The riskiness of escalating with audience costs has also prompted scholars to ask why leaders would “go public”, especially if doing so engages the audiences of the opponent. Kurizaki (2007) studies what happens if one were to relax the fifth assumption by giving actors the choice between public threats that create audience costs for both actors and private ones that do not. He finds that the risks inherent in the mechanism does prevent some threats from becoming public: leaders prefer to interact privately, where concessions might not be great but where it does not take much to satisfy an opponent who does not have to justify the outcome to a broader audience. Even Fearon’s (1994, 581) own evidence reveals such a dynamic: he notes that in 1755 British public troop movements made it harder for the French to back down, not just the British.

This brief discussion should make it clear that what often passes for ACT in empirical studies that investigate how regime type is linked to credibility, is in fact no theory at all. The question then is: does the evidence Trachtenberg presents help us advance the modeling enterprise?

2 What Do We Learn from the Cases?

Trachtenberg argues that (i) democratic leaders do not generate audience costs on purpose, (ii) even when they do, their opponents ignore them, and (iii) other factors better explain crisis outcomes.

Although Trachtenberg (2012, 6) discounts the importance of his first argument, it is the one he spends most of his time on and the one he presents most evidence for. I have no fundamental quarrel with that argument but one should realize just how indebted to ACT it is. Trachtenberg essentially claims that (democratic) leaders do not strategically attempt to generate audience costs because they do not want to be constrained by domestic opinion in foreign policy and because they fear the adverse effect explicit threats might have on the opponent or on third parties. Policy-makers evade audience costs by making vague threats whose plausible deniability enables them to obfuscate whether they have failed to follow through on any particular commitments, by finessing the outcomes to the point that it is unclear whether they have made any concessions at all, and by trying not to go public unless they are forced to. All of this means that audience costs should play little or no role at all provided leaders can get away with secrecy or are sufficiently able to mold public opinion.

From a purely logical perspective, this argument concedes more to ACT than some of the theoretical critiques mentioned above. Leaders can only be constrained by public opinion if making clear commitments and then backing down from them does generate audience costs. The premise of this argument is thus equivalent to the first assumption above. Moreover, its conclusion does not necessarily follow because leaders might not be able to deal in secret or manipulate public opinion. Trachtenberg provides evidence for both possibilities. For example, he notes that during the Fashoda crisis Salisbury was not a “free agent” and that he was forced to go public in order to show a belligerent public that he was not giving away the farm. Even though he did not deliberately seek to “stoke the public’s outrage”, the end effect was just the same and “even minor concessions could not now be made”
(14-15). Trachtenberg (2012, 8) also observes that even the autocratic Russian government could be “carried away by or unable to resist strong nationalist and pan-Slav feeling” during the Eastern Crisis of 1878, which caused them to refuse concessions until they found themselves isolated diplomatically and on the verge of war with Britain.

One case Trachtenberg does not discuss but that should have made the cut since it involves major powers of which one is a democracy, is the Sino-Indian War of 1962. The crisis preceding it is particularly revealing because Nehru had maintained secrecy for years in his border negotiations with the Chinese before the Opposition forced him to publish the letters. The public outcry in India was so strong, that the government dug in its heels and refused the concessions China was demanding, and even attempted to compel the militarily stronger opponent to capitulate with the “forward policy.” It is clear that China believed that Nehru had become committed to a policy of no concessions: they had to attack in order to show India the folly of such aggressive tactics. An issue that might have been amenable to a peaceful settlement in private became intractable once made public through the very mechanism ACT specifies.6

Trachtenberg (2012, 20) also provides direct evidence of leaders deliberately stoking public outrage, as Kinderlen did in the Second Moroccan Crisis of 1911. His threats did arouse nationalist feeling in Germany and the government “ended up paying a huge political price for what was seen as its willingness to accept a humiliating defeat”, as ACT assumes it would. Thus, we have evidence that leaders might not always avoid being constrained by their domestic audiences even if they wanted to, and we have evidence that sometimes they deliberately seek to inflame public opinion. Moreover, we have evidence that such constraints might occur in non-democracies as well, which is consistent with ACT as it stands now even if it does go contrary to the original “working hypothesis.”

The desire to avoid the loss of flexibility in foreign policy might not be the only reason policy-makers might be loath to make overt threats. Frightening the opponent might backfire if it raises his audience costs, provokes him into attacking, or gets third parties involved in the dispute to the detriment of the threatener. Trachtenberg provides evidence for all of these concerns, and they do constitute omitted and potentially distorting factors in the original ACT. But it is not like theorists are ignorant of this: Kurizaki (2007) explicitly deals with the possibility of a public threat increasing the cost of concessions for the opponent, and Slantchev (2010) shows that if a clear threat can provoke counter-measures that worsen the threatener’s military position, the threatener might avoid making such threats.

Thus, I agree with Trachtenberg, 7 that “public threat-making...is not necessarily an effective instrument of statecraft” even though he seems to be unaware of the work that deals with these issues, and at any rate fails to note just how dependent on ACT’s fundamental structure all of these arguments are.

Whatever the points of agreement might be, Trachtenberg (2012, 5) actually deemphasizes the first argument in favor of the second: even when (democratic) leaders go public – whether by design or not – adversaries do not infer that their hands are tied. Given the author’s own stress on this, it is very disappointing that it is precisely here that the paper is sorely lacking. When confronted with a case that might support ACT, Trachtenberg some-

6See Slantchev (2011, 178-90) for a case study of this crisis, and especially of the role public and military moves played in it ending with war.
times resorts to convoluted logic, interprets facts selectively, or altogether fails to provide any evidence at all.

Consider the Eastern Crisis, where Trachtenberg (2012, 10-11) is quick to dismiss the importance of the British government going public. We are told that the Disraeli government did not go public in order to limit his freedom of action, but to “make it clear... that Britain would go to war if Russia did not moderate her position.” This type of statement should really give one pause: even if we were to grant the unsubstantiated assertion that the British government was committed to war in the absence of concessions, we still have two critical questions. First, how much concessions would they need in order to be induced not to fight? Second, how were they going to signal this — i.e., “make it clear” — to the Russians? Since any concessions the Russians would agree to would be related to what they believed the British would be satisfied with (and thus depended on their estimate of the latter’s commitment to war), this must become a matter of signaling. It just will not do to assert, with scant evidence, that the Russians “did not doubt” the British commitment. What really matters is the extent of that commitment because, as Trachtenberg himself notes, Disraeli would have only resorted to force “if the terms were not satisfactory.” Thus, the British government did face the problem of convincing the Russians that they would have to give up more than they had originally believed. As to whether Disraeli’s threats were vague or not, I am unpersuaded by a sole reference to a speech Disraeli gave over a year before the crisis even began.

The only evidence for why British public statements did not matter for the Russians comes from the Russian ambassador Shuvalov who apparently advised his government not to be provoked by British bellicose statements and warlike public opinion (10). This is not the same as denying that this opinion might make it difficult for the British government to compromise. At any rate, Trachtenberg shows no evidence about what the Russian government thought of this. It might be instructive, however, that after the capitulation at Berlin, Shuvalov’s career was ruined because he was blamed for the Russian humiliation. It is also troubling that while Trachtenberg cites as evidence Shuvalov’s apparent lack of concern about British audience costs, he dismisses Gromyko’s concern about American audience costs in 1962. Recall that in the discussion of the Cuban Crisis of 1962, we are told that the Soviet foreign minister Gromyko warned that deploying the missiles in Cuba would create significant audience costs for the American President (33). Here we are asked to dismiss the importance of the public aspect of the crisis because Khrushchev apparently was not “particularly interested” in what he had to say. Setting aside the obvious concern that senior policy-makers apparently had with the possibility of the opponent being locked in by audience costs, in one instance we are asked to accept the confounding statement of one policy-maker while in another we are asked to dismiss the supportive one by another.

The Fashoda crisis is sometimes cited as supporting the original ACT although I do not believe that it does. One one hand, Trachtenberg and I seem to agree that the French were compelled by the British military moves, not by the public speeches or the publication of the Blue Book. On the other hand, Trachtenberg (2012, 16-17) goes too far in arguing that the Fashoda crisis “does not provide much support for a ‘rationalist explanation for war’ ” because “French policy in this affair is not to be understood in essentially rational terms.”

7 Slantchev (2011, 109).
But were the French truly so inept or deluded? It may have been true that it was known to
the French how Britain “felt about the subject” but this is by no means sufficient to deny
some rationale for the French policy, especially if we do not grant that France was bent on
having a “meaningful” presence in Egypt but instead wanted to end British occupation there
and perhaps secure some concessions elsewhere in Africa. The British were busy in Sudan,
the legality of their occupation of Egypt had been in dispute, and if France could count on
Russian support, an international conference might have secured some concessions along
the lines they would later obtain for Morocco. Trachtenberg dismisses any of this by arguing
that the British position was crystal-clear from the outset, and so was irrational to challenge.
Leaving aside the disturbing spaciousness of any explanation based on irrationality, the facts
that 477 members of parliament agreed to fund the French mission (only 18 opposed it), and
that the French capitulated once the British military moves signaled their commitment not
to concede at all suggests that the policy might have been a rational gamble that did have a
chance of working but simply did not pan out, in part because the Russian lack of support
was not easily foreseeable. In order to maintain the opposite, one must be prepared to accept
that 96% of French deputies were woefully inept or deluded, and even then one has to agree
that it was rational to scrap the attempt once the threat of war became clear.

What of the other cases Trachtenberg uses? In the Second Moroccan Crisis, he simply
provides no evidence at all that the public threats the British were making failed to influence
German calculations. In others, the theory itself does not seem to be applicable because
there were no military threats. For instance, in the First Moroccan Crisis, we are told that
France could not hope to deter Germany militarily, and thus had no interest in pursuing
tactics that would risk war. But the risk of commitment to war is the essential ingredient
of ACT — it is precisely how the threat is supposed to work as a coercive and signaling
device — and if one’s expected payoff from war is too low, then no such threats would be
made. The Berlin blockade reveals a similar problem because, as Trachtenberg, correctly
in my opinion, argues, neither side wanted to risk war. In situations like these, opting for
public warlike statements will be counter-productive, and one would expect states to pursue
strategies that shift the onus of escalation on the opponent, thereby trying to take advantage
of his reluctance to do anything that might provoke outright hostilities. One is hard-pressed
to see why one would even expect the mechanism of ACT to operate in these cases.

I was also somewhat surprised to read Trachtenberg’s (2012) interpretation of the 1945–
46 Iranian crisis. Although he does not tell us what he thinks the Soviets were after, he
seems to suggest that they were trying to conquer parts of northern Iran or otherwise expand
toward the Indian Ocean. This attempt is said to have been thwarted by vigorous Ameri-
can threats. But this was simply not so even if some American policy-makers believed it.
The evidence from the Russian archives suggests that destabilizing Iranian Azerbaijan (by
forming a local communist party in the northern area occupied by the Red Army, permitting
it to rebel, and then refusing to allow the Shah’s security forces to restore order) was not
an annexationist or expansionist move but a political ploy designed to wrest oil concessions
from Tehran. The Americans knew about the demand for oil concessions and although they
might have considered it legitimate, they certainly objected to the coercive methods the

---

8Fearon (1997); Slantchev (2011).
Russians were using.\textsuperscript{9} This support for the Iranians might have worked because the Russians moderated their demand by proposed the establishment of a Russo-Iranian joint-stock company (which they had previously refused) even if that meant abandoning the Azerbaijani communists. While the U.S. Ambassador regretted the “possibility that Iran will be forced to pay bribe to secure what should be accorded her automatically as of right,” he “did not feel that proposed solution is too bad.” The U.S. still urged Iran not to “purchase withdrawal of foreign troops” but Tehran thought they were running out of time.\textsuperscript{10} On April 4, the Iranians announced the agreement to the compromise (even though both its provisions were illegal under the law that Majlis had recently passed). They had negotiated secretly with the Soviets and had even misled the U.N. Security Council that no such negotiations were taking place.\textsuperscript{11} Having achieved their goal, the Soviets withdrew their troops. Even though the timing made it seem that they had done this under duress from the American public threats, the Soviet documents show that the decision was made before the row at the United Nations, and that the agreement “was of decisive importance.”\textsuperscript{12} It is true that the eventual outcome was an unmitigated disaster for the USSR: with the Red Army gone and the rebellion quashed, the newly-elected Majlis reneged on the deal, leaving the Soviets without concessions and with the public image of having been coerced by the Americans. However, the immediate outcome was distinctly favorable for the Russians, and this is why they withdrew, not because of American threats, which anyway came a little too late to be useful to the Iranians. The claim that these threats were effective is not supported by the evidence. Whether audience costs were relevant in making these threats more credible is simply beside the point. I do not know what, if anything, this case has to say about ACT and vice versa.

Trachtenberg’s (2012, 44) final argument is that a major reason ACT fails to account for crisis outcomes is that there are other factors that “play a much more important role”. Among those are overt military preparations, direct intelligence about the opponent’s intent from spies, and unexpected behavior by third parties. These are all sources of new information during the crisis and should certainly be expected to influence behavior. It would be just as strangely naive to assert that costly signaling is the only possible source of information as it would be to think that opponents could costlessly convey intent. What Trachtenberg fails to distinguish is that whereas new information can certainly be had, “indices” work precisely because they are beyond the control of the actor about whom they reveal something. Costly signaling theories, on the other hand, are concerned with how actors can make their behavior informative. Both have a role to play in crisis bargaining, and the most fruitful thing to do is to study how they interact.


\textsuperscript{11}Allen to the Secretary of State, May 11, 1946, FRUS 1946: 458.

\textsuperscript{12}Yegorova (1996, 17–19).
References


**URL:** [http://digital.library.wisc.edu/1711.dl/FRUS.FRUS1946v07](http://digital.library.wisc.edu/1711.dl/FRUS.FRUS1946v07)


