Response to Suzanne Werner and Amy Yuen "Making and Keeping Peace" ***International Organization*** **59:2 (Spring 2005), 261-292.**

The article by Werner and Yuen (hereafter W&Y) makes an important contribution to the still small, but growing, literature on the durability of peace. They argue that states sometimes return to war so as to renegotiate the terms of the settlement. For W&Y, two key variables affect the desire to renegotiate, and thus provide the key to understanding why some wars repeat while others yield lasting piece: 1) whether the war was interrupted by third-party pressure to reach a cease-fire, and 2) whether the military trend of the war was consistent so that it is clear how each side would fare if fighting continued.

W&Y set their argument up, in many ways, as a critique of my work – particularly my 2003 *IO* article "Scraps of Paper" and to a lesser extent, my 2004 book *Peace Time*. That my work provides a foil for theirs is very flattering, as it acknowledges my scholarship as the most important work in this field to date. However, I find their contribution and mine more complementary than contradictory. W&Y identify two important aspects of what I term "situational variables" that help shape the baseline prospects for peace. Their research coding battle consistency is a particularly valuable contribution that provides us with a better measure of the military outcome of the war, and therefore of the information that the fighting has revealed to each side about relative capabilities and resolve. Their coding of interrupted wars, similarly, provides an improved measure of incentives at the time the fighting stops. In other words, their research yields improved control variables for a study such as mine on the impact of the content of cease-fire agreements.

My basic finding that agreement strength matters holds up when these additional controls are added (W&Y table 1),[1] and agreement strength's contribution to overall model fit actually improves when these controls are added (W&Y p.281). This provides further evidence that controlling as much as possible for what I call the "baseline prospects for peace" is important for investigating the role of mechanisms deliberately chosen to increase the duration of peace. The better we can control for these baseline prospects, the clearer it is that agreements matter.

W&Y set the theoretical contrast between their work and mine as one between the distributional consequences of agreements and the enforcement aspects of agreements. They suggest that my work ignores the fact that one or both sides may actually prefer to go back to war to better the terms of settlement, emphasizing only that credible commitment problems and mistrust lead to renewed warfare. In this, they have mischaracterized my argument (in part, I believe, because they confuse and conflate it with Barbara Walter's work – this was more apparent in an earlier draft of their article which I commented on, but still persists to some degree).

I argue that there are several reasons that peace is fragile: the first of these is that recent belligerents have deeply conflicting interests and strong incentives to attack each other (see *Peace Time*, pp.13-15). In fact, my discussion of this obstacle to peace anticipates W&Y's

---

[1] Below, I address the contrary conclusion in part of W&Y's table 3.

argument about interrupted wars. I note (p.14) that "Reaching an acceptable settlement even after war may be particularly difficult if fighting ended in part because of international demands." Many of the mechanisms I focus on in my work are meant to alter or offset these incentives for aggression. In other words, "enforcement" aspects of agreements can ameliorate distributional problems, as well as problems of mistrust or the security dilemma. While analytically useful, the distinction between problems of aggression or a preference for war and fear is empirically difficult to make. Most recent belligerents are likely motivated by a combination of aggression and mistrust. While I assume that both sides prefer peace to war (as do the informational models of war on which W&Y base their work), I explicitly do not assume that they prefer peace on any terms to war (see *Peace Time* p.11). Thus when W&Y set their argument up as a corrective to the "general presumption" that maintaining peace is about reassurance, when a "more difficult problem" is that the belligerents may prefer "war to peace on the current terms" (W&Y p.164), they set themselves up against something of a straw-man.

Before turning to a discussion of some of the empirical issues raised by W&Y, one theoretical point is worth raising. W&Y base their argument on the idea that states fight again to alter the settlement between them. As they put it, settlements determine not just how peace will be enforced, but "who gets what and when" (pp.261-2). However, in the period examined in both their article and my work, wars surprisingly rarely settle the underlying political issues between the belligerents. This is an issue I will return to below, as it forms the puzzle that my new research project attempts to address. But if wars are no longer particularly useful for settling these political issues, then the theoretical premise of W&Y's argument is called into question.

W&Y raise a number of specific critiques of my empirical analysis. For example, they question the logic of dropping cases of joint democracy (p.276). In some of the statistical results reported in both my *IO* article and the book, I drop the few observations in which dyads become jointly democratic sometime after a cease-fire takes hold. Because there are so few instances that fit this description, it is not possible to control for the pacifying effects of joint democracy by including this variable in the model directly. Rather, I drop these observations as a way of controlling for this variable – the remaining variation takes place among the vast majority of cases that are not jointly democratic. As I note in the book (e.g, p.169), whether or not these few observations are dropped makes no substantive difference to the statistical results.[2]

W&Y also take issue with my treatment of changes in relative capabilities. Werner and I have something of a running debate about the role of this variable as a cause of the break down of peace. In a similar theoretical vein to W&Y, Werner's 1999 *AJPS* article argues that such changes are a primary cause of decisions by belligerents to go back to war so as to renegotiate settlement terms. In my work, I investigated such changes as a potentially important control

---

[2] Note also that because I employ time-varying covariates, each case consists of multiple observations. Dropping the joint democracy observations thus results in only parts of some cases being dropped. Only one full case is dropped from the analysis – the debatable case of Turkey and Cyprus in 1974. See *Peace Time* pp.110-111 for discussion.

variable. My criticism of this variable is not with the logic behind it, but with the empirical measure. Werner uses the COW annual capabilities index to create a measure of change in relative capabilities over time. Unfortunately, it is not clear from COW when during the year these measures are taken. This results in the possibility that changes in capabilities may occur after war breaks out rather than before, calling into question the direction of the causal arrow between changes and the resumption of war. The clearest example of this problem is the dramatic drop in the coding for Pakistan's capabilities in 1971 – Pakistan's CINC score drops from 0.0115 to 0.0087 from 1970 to 1971. This drop was clearly the result of Pakistan's loss of Bangladesh, which obviously affected its measure of population (which falls almost in half from 1970 to 1971), a key component of the CINC index. In other words the 1971 war changed the relative capabilities of Pakistan and India, not the other way around. This war took place in December, which implies that COW measures are taken at the end of the year (if the data were coded consistently, an admittedly big if). This means that a change in relative capabilities in a given year should not be used to explain the outbreak of war in that year, rather this measure should be lagged so that the purported cause clearly comes before the purported effect. As I note in my *IO* article, the positive effect of changes in relative capabilities on the risk of war (that is, the negative effect on peace) drops away when a lagged variable is used, casting doubt on Werner's finding that changes in capabilities cause war to resume (p.353).[3] In their *IO* article, W&Y maintain the practice of using the unlagged variable to test the effect of changes in capabilities. They argue that the 1971 war between India and Pakistan was caused by the attempted revolt and secession of Bangladesh – "the domestic chaos within Pakistan created a strategic opportunity for India to further weaken Pakistan" (p.277). That may be true, but a measure based on the COW capabilities index does not measure the effect of this domestic chaos, it measures population, energy consumption, iron and steel production, etc. Moreover, the successful secession of Bangladesh was caused by India's attack, not the other way around. In defense of Werner's argument about changes in capabilities, W&Y report (though they don't show these results) that over a longer data set (1816-1992), the lagged measure of relative capabilities has the hypothesized effect. This suggests to me that changes in relative capabilities do, in fact, have an effect on war recidivism, and that a lagged measure is appropriate for the reason given above. That W&Y do not use the lagged measure in their *IO* article thus seems highly problematic.

The India-Pakistan 1971 war provides a good example of the theoretical point raised above. While the 1971 war certainly weakened Pakistan dramatically – Pakistan's military defeat was overwhelming and it lost a significant portion of its territory – India did not press for a significantly better settlement over the Kashmir issue. India could easily have pushed for control of all of Kashmir as a condition of Pakistan's surrender, but chose not to. The "line of control" that became the dividing line in Kashmir in 1971 runs extremely close to the previous cease-fire line. (See *Peace Time* p.65 on this issue). To argue that India initiated this war to alter the

---

[3] Another problem with Werner's 1999 argument is that it is not clear, within the bargaining framework, why a change in relative capabilities that is known to both sides should cause war rather than a peaceful renegotiation of settlement terms.

settlement reached in the previous war (the Second Kashmir war) is thus untenable.

W&Y argue that their key variables, battle consistency and interrupted wars, have a greater substantive effect than does agreement strength. I have several responses to this claim. The first is that part of this difference is merely an artifact of the unit of measurement. To compare hazard ratios as a measure of the size of substantive effects across variables with different units of measurement is fairly meaningless. Interrupted war is a dummy variable, and battle consistency is measured on a 0 to 1 scale. Hazard ratios for those variables thus provide the increase or decrease in the risk of another war that takes place from one extreme of the variable to the other. In comparison, the index of agreement strength is measured on a scale from 0 to 10, so the hazard ratio provides the effect of just one tenth of the range of variation in this variable. Second, W&Y compare their results to only one of the measures of agreement strength I use in my work, not even mentioning (here, or anywhere in the article) the more subjective (but probably more accurate) measure of overall strength, or the effects of the specific mechanisms that go into the index of agreement strength. Third, and most important, while it is important to measure the size of substantive effects, my main interest is not in what factors have the largest effect on the dependent variable (on these grounds much more attention should be paid to variables such as whether the war ends in a tie and whether the war threatened the existence of either side). Rather, my interest is in whether, given the prospects for peace at the time of the cease-fire, anything can be done to make peace more likely to last. It is this question that has implications for theories of cooperation in international relations, and for policy-makers interested in affecting the chances for peace.

W&Y's section on alternative model specifications calls into question both the way I modeled the duration analysis and the significance of my finding about agreement strength. I chose to report the results of Weibull regressions rather than other specifications such as a Cox proportional hazards model for two reasons. First, as W&Y note, the Weibull provides more precise estimates in a small data set. Second, for many of the tests I report in both the IO article and the book, the Cox results are actually stronger than the Weibull results, so to be cautious I reported the set of results that were generally weaker.[4] It is true that for the regression that W&Y replicate in their article, the Cox results are slightly weaker. The significance level for the index of agreement strength in the Cox just barely misses the $p \leq .10$ cutoff ($p = .11$). As W&Y suggest,

[4] For example, the results (reported in table 2 in my *IO* article) for the subjective measure of agreement strength are generally stronger in the Cox model – the positive coefficients are larger for the "none" and "very weak" categories, and the individual significance levels of all but the "moderate" category are also higher (e.g., p drops from .29 to .02 for the "very weak" category). The results for many of the individual mechanisms are also stronger if a Cox model is used. For example, the results for armed peacekeepers (reported in table 4) are much stronger with the Cox, for new peacekeeping cases this result clears the $p<.001$ threshold (rather than the .10 threshold with the Weibull), similarly the size of the effect for joint commissions for dispute resolution is much higher in a Cox estimation (the coefficient goes from -16.69 as reported in table 5 to -39.70).

this difference between models could well be the result simply of the Cox model's tendency to overestimate standard errors in some cases.[5]

More troubling for my argument is W&Y's analysis of the shape of the hazard rate. As their figure 4 suggests, it may not be monotonic as the Weibull assumes. While it is not clear what model W&Y used to estimate the hazard rate for this figure making replication difficult, this issue warrants further investigation. Two points are worth raising for the time being, however. First, the results for the index of agreement strength appear to hold up fairly well in the log-logistic model that better approximates the shape of the hazard rate indicated in figure 4. Without W&Y's data, I have not replicated the results in table 3, model 4 that indicate this variable may no longer be significant, but the difference appears to be the result of larger standard errors introduced by adding variables rather than a drop in the substantive effect of agreement strength (the coefficient is the same in models 3 and 4). Second, and more interesting, is the question of why the hazard rate might rise at first and then fall, as indicated in W&Y's analysis. The shape of the hazard rate tells us something about whether peace gets harder or easier to maintain over time.[6] Intuitively, we would expect that peace is consolidated over time. This would also be consistent with an argument based on levels of mistrust – as the peace holds, the parties should become less mistrustful of each others intentions, making peace more likely to continue to hold. We should thus observe hazard rates falling over time. Informational models of war, such as the ones W&Y base their theory on, would predict the opposite. If war is caused by private information, and wars reveal this information credibly, then peace should be most stable in the immediate aftermath of war. Uncertainty about capabilities and resolve should build up again over time, making peace less stable. We should thus observe hazard rates rising over time after a cease-fire is in place. It is possible that these two dynamics are both at work, but this would seem more likely to yield a prediction that the hazard first fall and then rise, rather than the opposite. We might expect that the trust dynamic would dominate in the immediate aftermath of the fighting, but that the information dynamic would be more of a long-term process that would dominate later on. We should see thus see U-shaped hazard rate rather than a hill-shaped pattern. W&Y provide no theoretical explanation for why the hazard rate should take the shape that they suggest – the pattern they show in their article may simply be an artifact of their modeling decisions. This is clearly a substantive and theoretical issue (not just a methodological one) that warrants further investigation.

I have two small quibbles with the way W&Y construct their key variables. First, the

---

[5] In my book manuscript on peacekeeping in civil wars I also generally report Weibull estimates – both because of their added precision in relatively small data sets, and because the Weibull estimates for the effect of peacekeeping are consistently weaker than for the Cox, thus biasing results against my own argument. I do, however, include more explicit discussion of the Cox results, reporting some Cox estimates alongside Weibull estimates for comparison (see pp. 138-139 and tables 4.2 and 4.3 in *Peacekeeping and the Peackept*).

[6] As I discuss this in both *Peace Time* (pp.171-2) and the *IO* article (p.365).

interrupted war variable conflates two separate issues. One is whether the war ends with a cease-fire rather than a final settlement, the other is whether there is evidence of third-party pressure to end the fighting (p.273). It is not clear how they determined the first of these things – it is an issue that my variable "political settlement"(which they do not mention at all) attempts to measure. But their argument and the policy implications that they draw from it are both based on the latter. It seems to me that it would be more useful to analyze the effects of these concepts separately. A comparison of my political settlement variable and their interrupted war variable suggests that these two aspects may not be highly correlated. This raises questions about their causal logic, which contends that it is the third-party pressure that makes belligerents stop fighting before they reach a clear settlement.

Second, the explanation of the battle consistency measure is somewhat confusing. W&Y note (p.274) that they code cases of long-term stalemate as 0 for this variable (that is, these cases are coded as very inconsistent). This seems to conflate cases in which the military outcomes of individual battles vacillated widely (with one side doing very well, then the other) with those in which there was a long and consistent stalemate in which neither side could get the upper hand, even temporarily. The implications of these two types of cases should be very different for the duration of peace, according to their logic about revealed information. They suggest as much in the example they provide, the Korean war. They imply that the long-lived stalemate that was eventually reached in that war can account for the fact that peace has lasted to this day. This is exactly the opposite of their overall argument, that inconsistent battle outcomes (coded 0) should yield very unstable peace.

Several issues raised in W&Y point to interesting avenues for further research, some of which I am taking up in my new project on historical changes in war termination and the political decisiveness of war. As noted above, fundamental to the bargaining perspective on which W&Y base their analysis is the idea that wars yield political settlements. They begin their article with the argument that agreements "do not merely detail how the peace will be kept or specify the various enforcement mechanisms.... More fundamentally, agreements detail how the war ended – the terms of settlement.... [They] specify who gets what and when." (pp.261-2) Historically, this is an accurate description, but it no longer appears to hold true. Surprisingly, wars and the agreements that end them now rarely detail settlement of the political issues over which they were fought – they no longer specify who gets what and when. I note this in passing in *Peace Time* (pp.207-8). My new project will explore the causes and implications of this interesting change in the political decisiveness of war.

In their discussion of the effects of third-party intervention to end a war, W&Y focus on divergence between the political and the expected military outcome of the war. It is this divergence that gives belligerents an incentives to resume the fight. They argue that interruption by third parties can create this divergence, and further that such "pressure by third parties to ceasefire is increasingly commonplace" (p.269). They further note that such interruption is particularly likely when one side otherwise seems to be on the path to clear victory. This provides a potential hypothesis for why both military outcomes and the political decisiveness of wars may have changed in recent decades. I explore several related hypotheses in my paper on the decline of clear military victories. While it does seem to be true that third-party involvement **after** war (as peacekeepers) may be responsible for the rise of ties and the concomitant decline of military victories, patterns of external intervention **within** the war – which is closer to what W&Y mean by third-party interruption – cannot explain this change over time as patterns of intervention have not changed in a way that corresponds with the shift in outcomes. Nonetheless, W&Y's measure for "interruption" correlates highly with ties (see the correlation matrix on p.291) and it would be worth extending their measure back to wars before 1945 to explore this hypothesis further. A rise in wars interrupted by third party pressure also provides a potential explanation for the divorce between military and political outcomes. However, of the very few post-1945 interstate wars that were followed by an explicit agreement on the underlying issues (my "political settlement" variable), several are wars coded by W&Y as interrupted – including, most notably the Yom Kippur war, but also the Iran-Iraq war and the Football war. Nor do most of the uninterrupted wars yield political settlements. So it is not at all obvious that interruption prevents settlements (though it may distort them) or that uninterrupted war leads to a settlement that specifies "who gets what and when."

In sum, while I find W&Y's emphasis on the difference between our approaches somewhat overblown, and I take issue with several of the specific critiques of my own work, I think their article makes several important contributions. These include creating two new measures that help us better to control for the baseline prospects for peace in the study of post-war stability, and raising issues about political settlements emerging from war and the divergence between military and political outcomes that point in new and useful directions for research.