

PG 2
Scymade
Craig
Curt
Dave

Craig Haney

VIDEO TAPE ANALYSIS

(orig copy in
file)

"RA Report"

I. Methodological Aside

Undoubtedly the most complete record we have of the events which transpired last August, aside from our own recollections, is the video tape log. It is at the least remove from the reality of the prison of any of our data sources and, in terms of reinforcing the accuracy of our memory for the specifics of interaction and emotional tone for any event, the most valid and reliable since it is, unlike our own recall, uncluttered by post hoc interpretations. (In fact, as a veracity enhancing exercise, we might each view one or several hours of the video tapes just before writing the final draft of the study.)

Unfortunately, the processes by which we standardize the video record and transform events into statistical statements are more ambiguous than the record of the events themselves. What we would like to be able to say is: "This is the character of the events which constituted our prison." However, as a result of a number of serious distortions in the data, most of which derive from the vicissitudes of initial recording, our statements are rendered much more equivocal.

Even if our methods of data reduction, the translation of behavior into numbers, could have maintained perfect fidelity to the original, our video sample was seriously biased at the outset. For the most part, our concerns in filming were cinematic: we recorded primarily the dramatic

(e.g., counts) or the unusual (e.g., the priests' visit). In a sense, these events are unrepresentative by definition. It is the commonplace, the regular, the mundane occurrences which best portray the reality of a total institution and these are the very things we do not have (or have very sparsely) recorded.

In addition, because we used a novel format of behavioral scoring categories, we have no normative data from non-prison-simulation environments with which to compare our frequencies and correlations. Thus, all between-category comparisons (such as "there were more 'commands' than any other single behavioral event," or "four times as much 'aggression' as 'help' was observed to occur") must be interpreted at an absolute level since we simply cannot make statements relative to external situations. (The same must be said of the magnitude of the correlations between categories.) This is particularly unfortunate since it is precisely this kind of comparative information that allows us to represent the character of prison as distinct from, say, mental hospitals, educational settings, or cocktail parties.

Another way of stating this is to say that the variance in any behavioral category may derive from a variety of sources including: the nature of human nature or social interactions in general (i.e., any time people are behaving together they are likely to do these things), structural properties of the situations we selectively filmed (i.e., any time people are in certain situations together, like eating lunch for example, they are likely to do these things), definitional properties

of the categories themselves (i.e., one category is so much more broadly defined than others as to make its occurrence more probable in any situation) and, finally, properties specific to the prison environment itself. It is, of course, this latter in which we are most interested. But we simply do not have the necessary control groups to separate one source of variance from another. (Notice that even the absolute figures seem to carry comparative import deriving, I suspect from an implicit use of what might be called "intuitive control groups." That is, we say: in normal situations, "commands" do not constitute the most frequent behavioral event, nor is "aggression" four times more likely to occur than offers of "help." I think that our, as well as a reader's, intuitions are more likely to be correct on this than not -- yet we are still on empirically questionable ground without real normative data.)

II. The Scoring Categories

There were 25 relatively discrete incidents identifiable on the 6 tapes. Each incident or scene was scored for the presence of nine behavioral (including verbal) categories. These categories were defined as follows:

Question -- All questions asked, requests for information or assistance (excluding rhetorical questions, e.g., Korpi: "They don't take your clothes in a real prison. They do?")

Command -- An order to commence or abstain from a specific behavior, directed either to individuals or groups. Also generalized orders, e.g., "Settle down."

Information -- A specific piece of information proffered by anyone, whether requested or not, dealing with any contingency of the simulation.

Individuating Reference -- Positive: use of a person's real name, nickname, or allusion to special physical characteristics. Negative: use of prison number, title, or generalized "you."

Threat -- Verbal statement of contingent negative consequences of a wide variety, e.g., no meal, long count, pushups, lock-up in hole, no visitors, etc.

Deprecation/Insult -- Use of obscenity, slander, malicious statement directed toward individuals or group, e.g., "You lead a life of mendacity," or "You guys are really stupid."

Resistance -- Any physical resistance, usually prisoners to guards, such as holding onto beds, blocking doors, shoving guard or prisoner, taking off stacking caps, refusing to carry out orders.

Help -- Usually p → p or g → g; physically assisting another (i.e., excludes verbal statements of support). E.g., guard helping another to open door, prisoner helping another in cleanup duties.

Use of Instruments -- Use of any physical instrument to either intimidate, threaten, or achieve specific end. E.g., fire extinguisher, batons, whistles.

III. Frequencies

For each scene, the total number of behaviors in each category was summed, regardless of initiator or receiver. Thus, we have a total, for example, of all commands given, whether guard to prisoner, guard to guard,

staff to guard, etc. The means of these overall frequencies are presented below.

N = 25

	<u>Q</u>	<u>Comm</u>	<u>Inf</u>	<u>ID</u>	<u>Threat</u>	<u>Dep-I</u>	<u>Agg</u>	<u>R</u>	<u>H</u>	<u>I</u>
MEAN =	2.18	4.20	3.38	-1.36	1.24	2.02	1.02	1.04	.25	.94
SD =	2.12	3.38	2.00	2.16	1.75	2.33	1.62	1.27	.63	1.49

Without belaboring the point, keep in mind that the relative magnitudes of these numbers are partly the result of, for example, the scenes we chose to film. Thus, out of 25 separate incidents, almost half (11) were counts. This may very well explain the predominance of "Commands" and "Deprecation-Insult" in the overall frequencies.

IV. Frequency by Scene

These overall means can be broken down into category frequencies for each of the 25 scenes. Since scenes varied considerably in length, it was necessary to convert raw category frequencies into standard time scores so that scenes could be meaningfully compared for relative prevalence of specific behaviors. This was accomplished by dividing the frequency sum for each category by the length of the scene in question, as measured by tape units. The resulting number was multiplied by 100 to insure that primarily integers would be obtained.

CATEGORY FREQUENCIES

*(+) & (-) ID added

	<u>Scene & #</u>	<u>Q</u>	<u>Comm.</u>	<u>Info.</u>	<u>ID</u>	<u>Threat</u>	<u>Dep-I</u>	<u>Agg.</u>	<u>Res.</u>	<u>Help</u>	<u>Inst.</u>
Aug. 15	1. induction	.5	5	0	0	0	0	0	0	0	0
A	2. count	2.3	1.8	2.3	0	0	0	0	0	0	0
B	3. count -jw	0	13.5	0	0	0	.6	0	0	0	0
Aug. 16 C	4. count	0	1.4	0	0	0	.4	0	.7	0	.4
	5. breakfast	4	1	2	0	0	2	0	2	0	0
	6. rebellion	3.6	6.4	6.4	+2.1	0	3.6	3.6	3.6	0	3.6
A(+C)	7. count	2	12	4	-4	4	6	4	4	0	0
	8. lunch	2.5	1.3	5	-2.5	0	0	0	1.3	0	1.3
A	9. count	0	3	2.5	-2.5	1	1.5	.5	1	0	.5
	10. dinner	2.5	1.8	2.1	-1.7*	2.1	1.4	1.1	1.4	0	1.4
B	11. count - jw	4.2	6.3	2.1	-2.7*	3.7	6.8	.5	2.6	1.1	1.1
Aug. 17 C	12. count	1.1	3.5	.8	-1.6	.8	2.1	1.1	.5	.3	.8
A	13. count/work	.8	1.6	.8	-1.6	.8	0	0	0	.8	0
	14. lunch	2	2	2	0*	0	0	0	0	0	0
	15. dinner	4.3	1.4	2.9	-2.9	1.4	1.4	1.4	1.4	0	1.4
	16. visitors	.7	.3	2.3	-2.0*	.3	0	.3	0	.3	.3
	17. priest	6	6	3	-2	0	0	0	0	3	2
Aug. 18 A	18. count	1	5	4	-6	0	5	0	0	0	0
	19. induction	0	8	0	0	0	0	4	0	0	0
	20. dictation	0	4	8	0	0	4	0	0	0	0
	21. rebellion	.7	4.7	2.7	-1.3	4.7	1.3	4.7	3.3	.7	6.7
Aug. 19 A	22. count	.7	4	.7	-3.3	0	3.3	0	.7	0	0
	23. force eat	5.7	2.9	0	0	5.7	2.9	0	2.9	0	0
	24. visitors	8	0	4	0	4	0	0	0	0	2
B	25. count - jw	1.9	8.1	1.9	-7.5	2.5	8.1	4.4	.6	0	1.9

V. Directionality

Recall that the above frequencies represent simply occurrence of the category behavior, irrespective of initiator or recipient role. Usually finer grain analyses were not done on frequencies because directionality of some of the behaviors was for the most part obvious (as in the case of "Commands") and also because category frequencies would be so reduced as to make statistical analysis infeasible. For some of the categories, however, at least a percentage breakdown on the basis of directionality is informative.

<u>Questions:</u>	<u>Number</u>	<u>Percentage</u>
Prisoners of guards, staff	20	31.7
Staff, guards of prisoners	25	39.7
Others of guards, guards of others	7	11.1
Others of prisoners, prisoners of others	11	17.5
	<hr/> 63	<hr/> 100.0
 <u>Individuating Reference:</u>	 <u>Number</u>	 <u>Percentage</u>
Guards to prisoners	0	0
Prisoners to guards	3	100
	<hr/> 3	<hr/> 100

(Of more significance than the relative percentages here is the fact that only 3 individuating references between prisoners and guards were recorded.)

<u>De-individuating Reference:</u>	<u>Number</u>	<u>Percentage</u>
Guards to prisoners	43	74.1
Prisoners to guards	15	25.9
	<hr/> 58	<hr/> 100.0
 <u>Aggression:</u>		
Guards to prisoners	37	88.1
Prisoners to guards	5	11.9
	<hr/> 42	<hr/> 100.0
 <u>Resistance:</u>		
Guards to prisoners	2	5.9
Prisoners to guards	32	94.1
	<hr/> 34	<hr/> 100.0

VI. Count Data -- Frequencies

As mentioned earlier, the single event sampled most frequently on video tape was the count. It is perhaps worthwhile to abstract the behavior category frequencies for this event from the other scenes.

<u>Day</u>	<u>Scene</u>	<u>#</u>	<u>Q</u>	<u>Comm</u>	<u>Inf</u>	<u>ID</u>	<u>T</u>	<u>Dep-I</u>	<u>Agg</u>	<u>R</u>	<u>H</u>	<u>I</u>
Sun.	A	2	2.3	1.8	2.3	0	0	0	0	0	0	0
	B	3	0	13.5	0	0	0	0	0	0	0	0
Mon.	C	4	0	1.4	0	0	0	.4	0	.7	0	0
	A	7	2	12	4	-4	4	6	4	4	0	0
	A	9	0	3	2.5	-2.5	1	1.5	.5	1	0	.5
Tues.	B	11	4.2	6.3	2.1	-2.7	5.7	6.8	.5	2.6	1.1	1.1
	C	12	1.1	3.5	.8	-1.6	.8	2.1	1.1	.5	.3	.8
	A	13	.8	1.6	.8	-1.6	.8	0	0	0	.8	0
Wed.	C	18	1	5	4	-6	0	5	0	0	0	0
Thur.	A	22	.7	4	.7	-3.3	0	3.3	0	.7	0	0
	B	25	1.9	8.1	1.9	-7.5	2.5	8.1	4.4	.6	0	1.9
\bar{X}			1.27	5.47	1.74	-2.65	1.16	3.07	.95	.92	.20	.43
SD			1.21	3.96	1.35	2.35	1.46	2.82	1.57	1.21	.37	.59

In comparing the means for counts only with the overall means, note that, as we would expect, the frequencies of "Commands," "De-individuating References," and "Deprecation/Insult" are substantially higher. More interestingly, I think, levels of "Questions," "Information," and use of "Instruments" were much lower during the counts.

What happened during the counts is obviously important because these were the only formal periods of symbolic, sometimes actual, confrontation between the prisoners and guards. Higher levels of Comm, ID, and Dep-I suggest the guards were most definitely in control of the situation, making use of the obvious weapons available to them, giving orders and verbal abuse. Although the level of verbal behavior was probably much higher during the counts than at any other time, notice that the more civil forms of verbal interaction, namely asking questions and giving information, were actually less likely to occur. So not only was verbal aggression increased, but the normal, non-threatening modes of interaction were drastically reduced. (This is interesting because at the Air Force Academy we had a situation quite analogous to the counts: each evening, sometimes afternoons too, we would be lined up against the wall for counting off, instructions, and mostly general verbal abuse. A subtle, and the most effective, way we had of controlling the behavior of the upper classmen and thus co-opting even a little power, was by asking questions. A question is an excellent "foot-in-the-door" tactic because even in the most difficult situations it is likely to go unpunished and

be responded to since it tacitly reaffirms the other's superior power and knowledge -- even during periods of imposed silence, "Sir, may I ask a question?" was usually, even if inadvertently, permissible. Rarely did anyone think well enough in advance to categorically declare, "There will be no questions." Now, especially if it is asked in a reasonable, civil, humble manner, a question is difficult to respond to in an aggressive, hostile way. It usually results in a decent, if temporary, interchange. We practiced this with great skill at the Academy, not only decreasing the amount of time that could be used for verbal abuse, but also getting the upper classmen used to interacting with us in a more normal manner, as well as controlling their behavior by getting them to respond to us for a change. I took great pleasure in watching them get slowly infuriated but without knowing why. The same conditions applied during our prison but I never saw the tactic used. During the counts, when it can be used most effectively because of superior numbers, the level of question-asking actually went down.)

Also note that use of "Instruments" was much lower during the count. It would seem reasonable for this frequency to rise during the counts because here the superior numbers of the prisoners and the likelihood of being overpowered is most salient. Yet, the lack of instrument use here is really an effective symbolic demonstration on the part of the guards -- to control "all those prisoners" by purely verbal means is a much more dramatic show of strength, one which I am sure did not go unnoticed at some level by the prisoners.

VII. Count Analysis: Shift B (Eschleman) vs. Shifts A & C

One of the subjective impressions we (as well as the prisoners) developed was that John Wayne's shift was much more severe than others in its treatment of the prisoners. The video record of the counts provides an opportunity to test this and, in addition, determine the dimensions (in terms of our scoring categories) along which the shifts differed.

Shift B

Counts 2, 6, 11 (N=3)	<u>Q</u>	<u>Comm</u>	<u>Info</u>	<u>ID</u>	<u>Threat</u>	<u>De-I</u>	<u>Agg</u>	<u>R</u>	<u>H</u>	<u>I</u>
$\bar{X} =$	2.03	9.30	1.33	-3.40	2.07	5.17	1.63	1.07	.37	1.00
SD =	1.72	3.06	.95	3.10	1.54	3.27	1.97	1.11	.52	.78

Shifts A & C

Counts 1, 3, 4, 5, 6, 7, 8, 9, 10 (N=8)	<u>Q</u>	<u>Comm</u>	<u>Info</u>	<u>ID</u>	<u>Threat</u>	<u>Dep-I</u>	<u>Agg</u>	<u>R</u>	<u>H</u>	<u>I</u>
$\bar{X} =$.99	4.04	1.89	-2.38	.83	2.29	.70	.86	.14	.21
SD =	.77	3.23	1.45	1.90	1.26	1.80	1.30	1.24	.27	.30

t-tests of differences which were significant or near significant:

Commands $t = 2.50$ $p < .05$

Deprecation/Insult $t = 1.45$ $p < .20$

Instruments $t = 1.72$ $p < .20$

In addition to John Wayne's shift manifesting higher frequencies of "Commands," "Deprecation/Insult," and use of "Instruments," another interesting pattern emerges if we resort to a less conventional form of statistical comparison. Notice that over twice as many commands were issued,

over twice as many threats levied, over twice as many instances of deprecation/insult occurred, and over twice as many acts of aggression performed during John Wayne's shift. Yet the average frequency of prisoner "Resistance" is approximately the same. Thus, not only were prisoners receiving a much higher level of abuse from B shift, but they were much less likely to respond in kind to a hostile act with resistance. A strong measure of the control John Wayne and his shift exerted over the prisoners.

VIII. Count Analysis: Changes over Time

Another valuable source of data comes from looking at changes in category frequencies over time. Here we are asking how the character of the prison environment changed over time, e.g., was there a higher level of aggression at the beginning or the end of the study, were prisoners more likely to resist early in the study or after they had been prisoners for a few days, and so on. This kind of analysis is less subject to the methodological reservations raised at the outset because we are not trying directly to compare the prison environment with an external situation, but merely to detect what went on in the prison itself. (Here, too, I guess there is an untested but more reasonable assumption: that it is the passage of time in prison, that is, how long one has spent in the role of prisoner or guard, which is producing the changes in behavior, and not just the effects of time qua time.) Yet, some of the irregularities of data collection still complicate the strategy of analysis. Because we had free reign over which events to film, and because the nature

of the scene likely exercise a much greater effect on the behavior observed than any effect due to time, it is not clear what the unit of analysis in the T_1 vs. T_2 comparison should be. Fortunately, (and fortuitously), we did sample one event with enough regularity to make a before-after (or early-late) comparison meaningful. On the first (Sunday) and last full day (Thursday) of prison we recorded counts by guard shifts A and B, and these four counts may be used as an index of the ways in which the prison had changed in five days.

Shifts A & B

Counts 1 & 2 (N=2)	<u>Q</u>	<u>Comm</u>	<u>Info</u>	<u>ID</u>	<u>Threat</u>	<u>Dep-I</u>	<u>Agg</u>	<u>R</u>	<u>H</u>	<u>I</u>
$\bar{X} = 1.15$	7.65	1.15	0	0	.3	0	0	0	0	0
SD = 1.15	5.85	1.15	0	0	.3	0	0	0	0	0

Shifts A & B

Counts 11 & 12 (N=2)	<u>Q</u>	<u>Comm</u>	<u>Info</u>	<u>ID</u>	<u>Threat</u>	<u>Dep-I</u>	<u>Agg</u>	<u>R</u>	<u>H</u>	<u>I</u>
$\bar{X} = 1.30$	6.05	1.30	-5.40	1.25	5.70	2.20	.63	0	.95	
SD = .60	2.05	.60	2.10	1.25	2.40	2.20	.05	0	.95	

t-tests of differences:

(df = 2)

Use of ID $t = 3.65$ $p < .10$

Deprecation/Insult $t = 3.16$ $p < .10$

Resistance $t = 18.57$ $p < .01$

Because of the small number of degrees of freedom, statistical tests on these data are particularly stringent. It is worth noting that ID, Threat, Dep-I, Agg, R, and I did not occur at all (save trivial level of Dep-I) during the first two counts, and increased to fairly substantial

averages during the last two. With the exception of "Resistance" these categories all represent hostile abusive behaviors on the part of the guards and suggest that the character of the counts really did change quite dramatically over time. The level of "Commands" was actually greater in the initial counts but by the final two, all the dehumanizing modes of interaction had emerged full-blown. This is especially significant since one might expect the counts, being from the start the most formalized and stable form of interaction, to be less sensitive to these changes over time.

Master Copy

Video Summary

Prison Study
Craig Haney

1. Of the 10 categories of behavior scored, "Commands" had the highest frequency of occurrence (average of 4.2 per scene), followed by "Information" (3.38) and "Deprecation-Insult" (2.02).
2. Some of the percentage breakdowns for the directionality of some behavior are interesting:

Questions:

Prisoners to guards	-	31.7%
Guards to Prisoners	-	39.7%
Others	-	28.6%

Resistance:

Prisoners to guards	-	94.1%
Guards to Prisoners	-	5.9%

Aggression:

Prisoners to Guards	-	11.9%
Guards to Prisoners	-	88.1%

3. We have the necessary data for a comparison of Shifts A & B on counts 1 & 2 and 10 and 11 to index changes in behavior over time. This shows:

In counts 10 and 11, significantly more use of de-individuating references is made ($p < .10$)

($p < .10$) significantly, more use of deprecation-insult,

($p < .01$) and, significantly, more resistance occurs,

4. Finally, we have some "objective" support for the subjective impression that John Wayne's shift was more severe than the others, and also the behavioral dimensions along which that difference showed itself.

When B shift (John Wayne's) is compared against shifts A & C, it is found to have significantly more use of commands ($p < .05$), deprecation-insult ($p < .20$) and use of instruments ($p < .20$).